



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 138057**

**TO: James Schultz**  
**Location: rem/2d18/2c18**  
**Art Unit: 1635**  
**Monday, November 22, 2004**

**Case Serial Number: 10/001851**

**From: Alex Waclawiw**  
**Location: Biotech-Chem Library**  
**Rem 1A71**  
**Phone: 272-2534**

**[Alexandra.waclawiw@uspto.gov](mailto:Alexandra.waclawiw@uspto.gov)**

### **Search Notes**

This Page Blank (uspto)

STIC-Biotech/ChemLib

138057

mej

From: Schultz, James  
Sent: Wednesday, November 17, 2004 8:52 AM  
To: STIC-Biotech/ChemLib  
Subject: Seq Search 10/001,851

Hello,  
Please run a standard amino acid search on SEQ ID NO: 2 from the above entitled application. Instead of truncating the results at the top 15, could you please return the top 50 results from each database as well?  
Thanks,  
Doug Schultz

James Douglas Schultz, PhD  
AU 1635 (Biotechnology)  
Patent Examiner  
United States Patent and Trademark Office  
(Office) REM 2D18  
(Mail) REM 2C18  
(571) 272-0763

RECEIVED  
NOV 17 2004  
STIC-BIOTECH/ChemLib  
(STIC)

Point of Contact:

Alexandra Waclawiw

Technical Info. Specialist

\*\*\*\*\*  
STAFF USE ONLY

~~GM1-6A02 Tel: 308-449~~

Searcher: \_\_\_\_\_  
Searcher Phone: 2-\_\_\_\_\_  
Date Searcher Picked up: 11-22  
Date Completed: \_\_\_\_\_  
Searcher Prep/Rev. Time: 11-22  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search

NA Sequence: # \_\_\_\_\_  
AA Sequence: # (1)  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: November 22, 2004, 13:36:41 ; Search time 68 Seconds

(without alignments)  
3181.085 Million cell updates/sec

Title: US-10-001-851-2

3278

Sequence: 1 MRKKEKRLQAVLVLAALV.....TQQMLFPHNTSYLAEKRN 603

## Scoring table:

BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 35872929 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 100 summaries

## Database :

A\_Geneseq\_23Sep04:\*

1: geneseqp1980s:\*

2: geneseqp1980s:\*

3: geneseqp2000s:\*

4: geneseqp2000s:\*

5: geneseqp2002s:\*

6: geneseqp2003as:\*

7: geneseqp2003bs:\*

8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3278	100.0	603	AAU07778	AAU07778 Human nov
2	3278	100.0	603	AAU99098	AAU99098 Human gly
3	3272	99.8	603	ADG35387	ADG35387 Human N-a
4	3242.5	98.9	631	AAU07777	AAU07777 Human nov
5	2771	84.5	506	AAU07772	AAU07772 Human nov
6	2771	84.5	506	AAU07772	AAU07772 Human nov
7	2746.5	83.8	535	AAU07771	AAU07771 Human nov
8	1893	57.7	339	AAU94733	AAU94733 Human pro
9	1877	57.3	366	AAU07776	AAU07776 Human nov
10	1818	55.5	407	AAU07776	AAU07776 Human nov
11	1403.5	42.8	319	ABG08187	ABG08187 Human ace
12	1396	42.6	275	ADG25865	ADG25865 Human N-a
13	1392	42.5	276	ADG25865	ADG25865 Human N-a
14	1385.5	42.3	335	ABG08189	ABG08189 Human nov
15	1370	41.8	269	AAU07770	AAU07770 Human nov
16	1360	41.5	666	ABG08187	ABG08187 Human nov
17	1360	41.5	666	ABG08187	ABG08187 Human nov
18	1311	40.0	321	AAU07774	AAU07774 Human nov
19	1186	36.2	1252	ABG08190	ABG08190 Human nov
20	1155	35.2	499	ABG08190	ABG08190 Human nov
21	1125	34.3	559	ABG08190	ABG08190 Human nov
22	1125	34.3	561	ABG08190	ABG08190 Human nov
23	1116	34.0	559	ABG08190	ABG08190 Human nov
24	1116	34.0	559	ABG08190	ABG08190 Human nov
25	1116	34.0	559	ABG08190	ABG08190 Human nov

26	1115	34.0	517	AA66402	AA66402 GalNac-tr
27	1115	34.0	517	AA66402	AA66402 Honeybee
28	1115	34.0	559	AA66401	AA66401 Cattle Ga
29	1115	34.0	559	AA66401	AA66401 Cattle Ga
30	1115	34.0	559	AA66401	AA66401 Cattle Ga
31	1113.5	34.0	630	ABG64271	ABG64271 Bovine N-
32	1112	33.9	571	ADG10116	ADG10116 Drosophi
33	1106	33.7	556	ABG32379	ABG32379 Human NOV
34	1106	33.7	556	ABG32379	ABG32379 Human NOV
35	1106	33.7	556	ABG32379	ABG32379 Human NOV
36	1096.5	33.5	555	ADG10114	ADG10114 Human NOV
37	1085.5	33.1	573	ADG04927	ADG04927 Sea equi
38	1050	32.0	657	AA888419	AA888419 Human mem
39	1048	32.0	478	ABR41518	ABR41518 Human DIT
40	1048	32.0	478	ADL22696	ADL22696 Human dis
41	1045.5	31.9	658	AAW78711	AAW78711 Human pro
42	1040	31.7	930	ADG5070	ADG5070 Rat Prote
43	1035.5	31.6	591	ABG62319	ABG62319 Drosophi
44	1026	31.3	578	ABU07527	ABU07527 Human N-a
45	1024.5	31.3	188	ADG25974	ADG25974 N-acetyl
46	1015	31.0	610	ADG5072	ADG5072 Human pro
47	1013	30.9	940	AAV53040	AAV53040 Human sec
48	1013	30.9	940	ADK70539	ADK70539 Respirato
49	999.5	30.5	558	ABP69229	ABP69229 Human pol
50	999.5	30.5	558	AAE25021	AAE25021 Human dru
51	991.5	30.2	581	ABG32510	ABG32510 Human gly
52	991.5	30.2	581	AAE25019	AAE25019 Human dru
53	991.5	30.2	581	ADH48838	ADH48838 NOV52 pro
54	990	30.2	240	AAU07773	AAU07773 Human nov
55	988	30.1	209	AAU07773	AAU07773 Human nov
56	986.5	30.1	581	ADG35388	ADG35388 Human N-a
57	986.5	30.1	581	ADG35388	ADG35388 Human N-a
58	977.5	29.8	560	ADG25922	ADG25922 Human pro
59	971.5	29.6	581	ADG25914	ADG25914 Novel N-a
60	971.5	29.6	552	AAE06620	AAE06620 Human pro
61	971	29.6	552	AAU12269	AAU12269 Human pro
62	971	29.6	552	AAU09026	AAU09026 Human nov
63	971	29.6	552	ABO17713	ABO17713 Novel hum
64	971	29.6	552	ABR58585	ABR58585 Human can
65	971	29.6	552	ABU080967	ABU080967 Human pro
66	971	29.6	552	ABU66667	ABU66667 Human pro
67	971	29.6	552	ABU59748	ABU59748 Novel sec
68	971	29.6	552	ABO24938	ABO24938 Human sec
69	971	29.6	552	ABU66943	ABU66943 Human sec
70	971	29.6	552	ADA45715	ADA45715 Novel hum
71	971	29.6	552	ADA76146	ADA76146 Human pro
72	971	29.6	552	ADA18796	ADA18796 Human pro
73	971	29.6	552	ADA61419	ADA61419 Homo sapi
74	971	29.6	552	ADA19204	ADA19204 Novel hum
75	971	29.6	552	ADG27745	ADG27745 Human pro
76	971	29.6	552	ADA66224	ADA66224 Novel hum
77	971	29.6	552	ADA15788	ADA15788 Human pro
78	971	29.6	552	ADA47574	ADA47574 Human pro
79	971	29.6	552	ADA67369	ADA67369 Human pro
80	971	29.6	552	ADG30376	ADG30376 Human pro
81	971	29.6	552	ADA85672	ADA85672 Human pro
82	971	29.6	552	ADA96884	ADA96884 Human pro
83	971	29.6	552	ADA79188	ADA79188 Human pro
84	971	29.6	552	ADA87327	ADA87327 Novel hum
85	971	29.6	552	ADA16529	ADA16529 Human pro
86	971	29.6	552	ADA16529	ADA16529 Human pro
87	971	29.6	552	ADA16529	ADA16529 Human pro
88	971	29.6	552	ADA16529	ADA16529 Human pro
89	971	29.6	552	ADA16529	ADA16529 Human pro
90	971	29.6	552	ADA16529	ADA16529 Human pro
91	971	29.6	552	ADA16529	ADA16529 Human pro
92	971	29.6	552	ADA16529	ADA16529 Human pro
93	971	29.6	552	ADA16529	ADA16529 Human pro
94	971	29.6	552	ADA16529	ADA16529 Human pro
95	971	29.6	552	ADA16529	ADA16529 Human pro
96	971	29.6	552	ADA16529	ADA16529 Human pro
97	971	29.6	552	ADA16529	ADA16529 Human pro
98	971	29.6	552	ADA16529	ADA16529 Human pro

99 971 29.6 552 6 ADB29824  
100 971 29.6 552 6 ADA80352

## ALIGNMENTS

## RESULT 1

AAU07778  
ID AAU07778 standard; protein; 603 AA.

AC AAU07778;

DT 04-DEC-2001 (first entry)

DE Human novel transferase protein, NHP #21.

KM Human; transferase; breast cancer; prostate cancer; immunogen;

KW gene therapy; antisense.

OS Homo sapiens.

PN WC020164903-A2.

PD 07-SEP-2001.

PF 28-FEB-2001; 2001WO-US006460.

PR 29-FEB-2000; 2000US-0185920P.

PR 02-MAR-2000; 2000US-0186558P.

PR 24-MAR-2000; 2000US-0191849P.

XX (LEXI-) LEXICON GENETICS INC.

PI Donoho G, Hilbun E, Turner CA, Friedrich G, Abuin A;

PI Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kleke JA;

PI Potter DG;

XX WPI; 2001-550185/61.

XX N-PSDB; AAS12625.

XX Novel nucleic acid sequences encoding novel human proteins useful for

XX diagnosis, drug screening, clinical trial monitoring and treatment of

XX diseases and disorders.

XX Claim 5; Page 56-58; 60pp; English.

XX The invention relates to isolated nucleic acids encoding novel human

XX transferase proteins (NHP). The nucleic acids and proteins are useful for

XX diagnosis, drug screening, clinical trial monitoring and treatment of

XX diseases and disorders e.g. breast and prostate cancer. NHPs can also be

XX useful for augmenting the efficacy of chemotherapeutic agents used in

XX treatment of breast or prostate cancer. The nucleic acid is also useful

XX in NHP gene regulation, and as antisense primers in amplification

XX reactions of NHP gene sequences. NHPs are useful for producing

XX antibodies. The present sequence represents a novel human transferase

XX which has sequence similarity to N-acetyl-galactosaminyltransferase

XX Sequence 603 AA;

Query Match 100.0%; Score 3278; DB 4; Length 603;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRKKRLQAVVLAVALVLLPNYGLMALYRERPDGPGSGAAVAPAAQSGSHSROK 60

DB 1 MRRKKRLQAVVLAVALVLLPNYGLMALYRERPDGPGSGAAVAPAAQSGSHSROK 60

QY 61 KTFPLGDOQKLDMDKKAIRRDQVNGGEGRPYMTDARVNDQVARENGFNLYVSDK 120

DB 61 KTFPLGDOQKLDMDKKAIRRDQVNGGEGRPYMTDARVNDQVARENGFNLYVSDK 120

QY 121 ISLRSLPDIRHPNCSKRYLETLPNTSIIIPFHNESGSLLRVTAVLSNPPELVAEI 180

DB 121 ISLRSLPDIRHPNCSKRYLETLPNTSIIIPFHNESGSLLRVTAVLSNPPELVAEI 180

QY 181 VLVDPSDREHLKCPLEDYVALFPSVRLTPKKEGLIRPMIGASVATGVITFLSHC 240

DB 181 VLVDPSDREHLKCPLEDYVALFPSVRLTPKKEGLIRPMIGASVATGVITFLSHC 240

QY 241 EAVNMLPPLDRIARNRKTIVCPMIDVIDHDDEFRYEQADANRGAFDMEYKRLPI 300

DB 241 EAVNMLPPLDRIARNRKTIVCPMIDVIDHDDEFRYEQADANRGAFDMEYKRLPI 300

QY 301 PELQADPSDFESPVNAGLFAVDRKMFELGYPDGLIWSGEQYEISFKVMCGGRM 360

DB 301 PELQADPSDFESPVNAGLFAVDRKMFELGYPDGLIWSGEQYEISFKVMCGGRM 360

QY 361 EDIPCSRVGHYRKYVYKYPAGVSLARNLKRVAEVMWDEYAEIYQRRPEYRLSAGDV 420

DB 361 EDIPCSRVGHYRKYVYKYPAGVSLARNLKRVAEVMWDEYAEIYQRRPEYRLSAGDV 420

QY 421 AVOKKLRSSLNCKSFKPMFTKIAMDLPKFYPPVPPAAAMGIRNVGTGLCADTKHGLG 480

DB 421 AVOKKLRSSLNCKSFKPMFTKIAMDLPKFYPPVPPAAAMGIRNVGTGLCADTKHGLG 480

QY 481 SPLRLGCGVCRGSAANNQVFTFTWRBDIRPDPOHTKKCFDAISHTSPVTLYDCHS 540

DB 481 SPLRLGCGVCRGSAANNQVFTFTWRBDIRPDPOHTKKCFDAISHTSPVTLYDCHS 540

QY 541 MKGNQMKYRKDKTLHPVSGSCMDSCSDHRIIPMTCPBSSLTQOMLFHTNSTVLEKF 600

DB 541 MKGNQMKYRKDKTLHPVSGSCMDSCSDHRIIPMTCPBSSLTQOMLFHTNSTVLEKF 600

QY 601 NRN 603

DB 601 NRN 603

## RESULT 2

AAU99098  
ID AAU99098 standard; protein; 603 AA.

XX AAU99098;

XX 30-AUG-2002 (first entry)

XX Human glycosyl transferase 47169.

XX Human; glycosyl transferase; 47169; 33935; cancer; carbohydrate storage;

XX diabetes mellitus; hypoglycaemia; arthritis; rheumatism;

XX autoimmune disorder; systemic lupus erythematosus; Grave's disease;

XX myasthenia gravis; insulin resistance; scleroderma; rheumatoid arthritis;

XX autoimmune infertility; tumorigenesis.

XX Homo sapiens.

XX WO200240657-A2.

XX 23-MAY-2002.

XX 20-NOV-2001; 2001WO-US047575.

XX 20-NOV-2000; 2000US-0249939P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Meyer R, Williamson M;

XX WPI; 2002-508326/54.

XX N-PSDB; ABK86093.

XX Use of modulator of activity of novel glycosyl transferase proteins,

XX 47169/33935 proteins, for making a medicament for modulating ability of

XX cell to affect glycosylation state of lipid or polypeptide target in

XX  
PS Claim 27; Fig 1; 153pp; English.

CC The invention relates to the use of a modulator of the activity of an  
CC enzyme such as 47169 or 33935 protein (a novel glycosyl transferase  
CC protein) for making a medicament for modulating the ability of a cell to  
CC affect the glycosylation state of a lipid target or polypeptide target in  
CC a cell. Also included is a method of assessing (M1) if a test compound is  
CC useful for modulating at least one phenomenon (P) such as non-covalent  
CC binding between a protein and one of a cell, a virus and another protein;  
CC cell signalling, cell differentiation, tumorigenesis, cell adhesion, cell  
CC motility, cell-to-cell interaction, cell invasiveness, cell proliferation,  
CC gene transcription, and an immune response, comprising: (a) adding the  
CC test compound to a first composition comprising a 603 residue 47169  
CC polypeptide sequence (S2), or a sequence at least 90 % identical to a 492  
CC residue 33935 polypeptide sequence (S12), both given in the specification  
CC ; and (b) comparing the activity in the first composition and in a second  
CC composition that is substantially identical to the first composition,  
CC except that it lacks the test compound, whereby a difference in the  
CC activity in the first and second compositions is an indication that the  
CC test compound is useful for modulating the phenomenon. The method is  
CC useful for making a medicament for modulating the ability of a cell (e.g.  
CC human endothelial cell such as lung cell, breast cell or colon cell,  
CC preferably a tumour cell) to affect the glycosylation state of a target  
CC such as lipid or polypeptide. The identified test compound is useful for  
CC treating disorders such as diabetes mellitus, hypoglycaemia, arthritis,  
CC rheumatism, autoimmune disorders (e.g. systemic lupus erythematosus,  
CC Grave's disease, myasthenia gravis, insulin resistance, rheumatoid  
CC arthritis, scleroderma and autoimmune infertility), tumorigenesis,  
CC cancer and tumour metastasis. The present sequence represents glycosyl  
CC transferase 47169

CC  
XX Sequence 603 AA;

Query Match 100.0%; Score 3278; DB 5; Length 603;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRERKRLQAVLVLAALVLLPVGIMALYRERQPGTGGGGAAPAAAGGSHSRK 60  
DB 1 MRRERKRLQAVLVLAALVLLPVGIMALYRERQPGTGGGGAAPAAAGGSHSRK 60  
QY 61 KTFPLDGGGKDKMHDEKARLDAQRVNGEQRPYMTDAERVDQAYRENGFIIVYSDK 120  
DB 61 KTFPLDGGGKDKMHDEKARLDAQRVNGEQRPYMTDAERVDQAYRENGFIIVYSDK 120  
QY 121 ISLNRSLPDIRHPVCKNSKRYLETLPNTSIIIPFNEGMSSILRTVHSLVLRSPPELVAEI 180  
DB 121 ISLNRSLPDIRHPVCKNSKRYLETLPNTSIIIPFNEGMSSILRTVHSLVLRSPPELVAEI 180  
QY 121 ISLNRSLPDIRHPVCKNSKRYLETLPNTSIIIPFNEGMSSILRTVHSLVLRSPPELVAEI 180  
DB 121 ISLNRSLPDIRHPVCKNSKRYLETLPNTSIIIPFNEGMSSILRTVHSLVLRSPPELVAEI 180  
QY 181 VLVDPSDRHLKKPLEDYALPPSVRIILRTKKEGLIRTRMLGASVATGDIITFLDSHC 240  
DB 181 VLVDPSDRHLKKPLEDYALPPSVRIILRTKKEGLIRTRMLGASVATGDIITFLDSHC 240  
QY 241 EAVNNWMLPILLDIRAARAKTIVCMIDVIDDHFRTYQAGDARQAFDMWMTYKRIPIR 300  
DB 241 EAVNNWMLPILLDIRAARAKTIVCMIDVIDDHFRTYQAGDARQAFDMWMTYKRIPIR 300  
QY 301 PELQKADPSDFESPVAAGGLFAVDRKFMELAGYDGLETIWMGEQYEISPKVMCCGRM 360  
DB 301 PELQKADPSDFESPVAAGGLFAVDRKFMELAGYDGLETIWMGEQYEISPKVMCCGRM 360  
QY 361 EDIPCSFVGHIIYRKYVPYKPVAGVSLARNLKRVAVEMDEYAEIYORRPEYRHLASGV 420  
DB 361 EDIPCSFVGHIIYRKYVPYKPVAGVSLARNLKRVAVEMDEYAEIYORRPEYRHLASGV 420  
QY 421 AVQKGLSSLNCKSFKKFMKTIAMDLPKFPYVPPVPAAMWEIRNNGTGLCAQDKHAGLG 480  
DB 421 AVQKGLSSLNCKSFKKFMKTIAMDLPKFPYVPPVPAAMWEIRNNGTGLCAQDKHAGLG 480  
QY 481 SPLRLBECVGRGGAANNMNVQVFTFTWRERDIPGDPHTTKFCFDAISHTSPVTLVYCHS 540  
DB 481 SPLRLBECVGRGGAANNMNVQVFTFTWRERDIPGDPHTTKFCFDAISHTSPVTLVYCHS 540

QY 541 MKGNQLMKRYKDKTLYHPVSGCMDCSESDHRIEMNTCNPSLTOQLFEHTNSTYLEKF 600  
DB 541 MKGNQLMKRYKDKTLYHPVSGCMDCSESDHRIEMNTCNPSLTOQLFEHTNSTYLEKF 600  
QY 601 NRN 603  
DB 601 NRN 603

RESULT 3

ID ADC35387 standard; protein; 603 AA.

XX ADC35387;

DT 18-DEC-2003 (first entry)

XX Human N-acetylglucosamine transferase (GalNAc) T13 protein.

XX N-acetylglucosamine transferase; GalNAc; alpha1-bond;

XX genetic engineering; GalNAc-T13; cytosolic; cancer treatment;

XX cancer diagnosis; gene therapy; human; enzyme.

XX Homo sapiens.

XX WO2003057897-A1.

XX 17-JUL-2003.

XX 06-JAN-2003; 2003WO-JP000008.

XX 28-DEC-2001; 2001JP-00401507.

XX 05-JUN-2002; 2002JP-00163832.

XX 12-JUL-2002; 2002JP-00203696.

XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

XX (AMSH) AMERSHAM BIOSCIENCES KK.

XX (FURE) FUJIREBIO INC.

XX Narimatsu H, Zhang Y, Gotoh M;

XX WPI; 2003-567133/55.

XX N-PSDB; ADC35394.

XX Novel UDP-N-acetyl-D-galactosamine:polypeptide N-acetylglucosamine

XX transferase and encoded nucleic acid, applicable in identifying O-bonded

XX sugar-attached proteins, gene diagnosis, and cancer therapy.

XX Claim 1; SEQ ID NO 3; 153pp; Japanese.

XX This invention relates to a novel protein with an activity of

XX transferring N-acetylglucosamine (GalNAc) via an alpha1-bond to a

XX hydroxyl group of serine or threonine in a protein or a peptide sequence.

XX The invention provides gene and protein sequences for these enzymes to

XX enable genetic engineering or production of these enzymes. Seven genes

XX and their proteins are described by the invention, GalNAc-T13 to T17. The

XX enzymes of the invention may have cytosolic activity, and hence may be

XX used for treatment or diagnosis of cancer, and the sequences described

XX may be useful in gene therapy. The present sequence is the partial

XX sequence of the human GalNAc-T13 transferase protein of the invention.

XX Sequence 603 AA;

Query Match 99.8%; Score 3272; DB 7; Length 603;

Best Local Similarity 99.8%; Pred. No. 0;  
Matches 602; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRERKRLQAVLVLAALVLLPVGIMALYRERQPGTGGGGAAPAAAGGSHSRK 60  
DB 1 MRRERKRLQAVLVLAALVLLPVGIMALYRERQPGTGGGGAAPAAAGGSHSRK 60  
QY 61 KTFPLDGGGKDKMHDEKARLDAQRVNGEQRPYMTDAERVDQAYRENGFIIVYSDK 120

Db 61 KTFPLGSGQKLDHMDKEAIRDAQRVNGEGRPYPMTDARVDQAVRENGFNIVYSDK 120  
Qy 121 ISLRSLLPDIHHPNCSKRYETLPTNTSIIIPFNHEGSSLLRTVHSLNRPPELVAEI 180  
Db 121 ISLRSLLPDIHHPNCSKRYETLPTNTSIIIPFNHEGSSLLRTVHSLNRPPELVAEI 180  
Qy 181 VLVDPSDRHLKKPLBDYMLFPSVRLRTKKEGLRTMTLGSVAATGVITFLDSHC 240  
Db 181 VLVDPSDRHLKKPLBDYMLFPSVRLRTKKEGLRTMTLGSVAATGVITFLDSHC 240  
Qy 241 EANNVMLPPLIDRLIARRKTIIVCPMIDVIDHDDFRYETQAGDAMRGAFDMEMTYKRIPI 300  
Db 241 EANNVMLPPLIDRLIARRKTIIVCPMIDVIDHDDFRYETQAGDAMRGAFDMEMTYKRIPI 300  
Qy 301 PELQKADSPDPESPVNAGGLFAVDKRMFWELGYDPLGLTWGGEQYEISFKVMWCGGRM 360  
Db 301 PELQKADSPDPESPVNAGGLFAVDKRMFWELGYDPLGLTWGGEQYEISFKVMWCGGRM 360  
Qy 361 EDICSRVGHYRKVPYKVPAGVSLAARLKRVAEVMDEYAEVYQRRPEYRHLASGDV 420  
Db 361 EDICSRVGHYRKVPYKVPAGVSLAARLKRVAEVMDEYAEVYQRRPEYRHLASGDV 420  
Qy 421 AVQKKLRSSLNCKSFKMTKIAMDLPRFYPPVEPPAAWGEIRNVGTGLCADTYHGALG 480  
Db 421 AVQKKLRSSLNCKSFKMTKIAMDLPRFYPPVEPPAAWGEIRNVGTGLCADTYHGALG 480  
Qy 481 SPLLEGCVRGGEAANNMOVFTFTWREDIRPGDPQHTKKRCPAISHTSPTVTLYDCHS 540  
Db 481 SPLLEGCVRGGEAANNMOVFTFTWREDIRPGDPQHTKKRCPAISHTSPTVTLYDCHS 540  
Qy 541 MKGNQLMKRYKDKTLHYHVSQSCMDSCSDHRIFNATCNPSLSLTQOMLFERTNSTVLEKF 600  
Db 541 MKGNQLMKRYKDKTLHYHVSQSCMDSCSDHRIFNATCNPSLSLTQOMLFERTNSTVLEKF 600  
Qy 601 NRN 603  
Db 601 NRN 603

RESULT 4  
AAU07777  
ID AAU07777 standard; protein; 631 AA.  
XX  
AC AAU07777;  
XX  
DT 04-DEC-2001 (first entry)  
XX  
DE Human novel transferase protein, NHP #20.  
XX  
KM Human; transferase; breast cancer; prostate cancer; immunogen;  
XX  
KM gene therapy; antisense.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 630..631  
XX /note= "Encoded by AATGSAAC"  
XX  
XX WO200164903-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 28-FEB-2001; 2001MO-US006460.  
XX  
PK 29-FEB-2000; 2000US-0185920P.  
PR 02-MAR-2000; 2000US-018558P.  
PR 24-MAR-2000; 2000US-0191849P.  
XX  
PA (LEXI-) LEXICON GENETICS INC.  
XX  
XX Donoho G, Hilbun E, Turner CA, Friedrich G, Abuin A,  
PI Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kleke JA;

PI Potter DG;  
XX WPI; 2001-550185/61.  
XX N-P6DB; AAS12624.  
XX  
XX Novel nucleic acid sequences encoding novel human proteins useful for  
PT diagnosis, drug screening, clinical trial monitoring and treatment of  
PT diseases and disorders.  
XX  
XX Claim 6; Page 54-55; 60pp; English.  
XX  
XX The invention relates to isolated nucleic acids encoding novel human  
CC transferase proteins (NHP). The nucleic acids and proteins are useful for  
CC diagnosis, drug screening, clinical trial monitoring and treatment of  
CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be  
CC useful for augmenting the efficacy of chemotherapeutic agents used in  
CC treatment of breast or prostate cancer. The nucleic acid is also useful  
CC in NHP gene regulation, and as antisense primers in amplification  
CC reactions of NHP gene sequences. NHPs are useful for producing  
CC antibodies. The present sequence represents a novel human transferase  
CC which has sequence similarity to N-acetyl-galactosaminyltransferase  
SQ  
Sequence 631 AA;  
Query Match 98.9%; Score 3242.5; DB 4; Length 631;  
Best Local Similarity 95.4%; Pred. No. 0;  
Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;  
Qy 1 MRKREKLLQVALVALVALVLLPVGGLMALYREOPDPTGGSGAAVAPAGGSHSRQK 60  
Db 1 MRKREKLLQVALVALVALVLLPVGGLMALYREOPDPTGGSGAAVAPAGGSHSRQK 60  
Qy 61 KTFPLGSGQKLDHMDKEAIRDAQRVNGEGRPYPMTDARVDQAVRENGFNIVYSDK 120  
Db 61 KTFPLGSGQKLDHMDKEAIRDAQRVNGEGRPYPMTDARVDQAVRENGFNIVYSDK 120  
Qy 121 ISLRSLLPDIHHPNCSKRYETLPTNTSIIIPFNHEGSSLLRTVHSLNRPPELVAEI 180  
Db 121 ISLRSLLPDIHHPNCSKRYETLPTNTSIIIPFNHEGSSLLRTVHSLNRPPELVAEI 180  
Qy 181 VLVDPSDRHLKKPLBDYMLFPSVRLRTKKEGLRTMTLGSVAATGVITFLDSHC 240  
Db 181 VLVDPSDRHLKKPLBDYMLFPSVRLRTKKEGLRTMTLGSVAATGVITFLDSHC 240  
Qy 241 EANNVMLPPLIDRLIARRKTIIVCPMIDVIDHDDFRYETQAGDAMRGAFDMEMTYKRIPI 300  
Db 241 EANNVMLPPLIDRLIARRKTIIVCPMIDVIDHDDFRYETQAGDAMRGAFDMEMTYKRIPI 300  
Qy 301 PELQKADSPDPESPVNAGGLFAVDKRMFWELGYDPLGLTWGGEQYEISFK----- 352  
Db 301 PELQKADSPDPESPVNAGGLFAVDKRMFWELGYDPLGLTWGGEQYEISFK----- 352  
Qy 353 -----VNMCGRMEDIPCSRGHYRKVPYKVPAGVSLAARLK 391  
Db 353 -----VNMCGRMEDIPCSRGHYRKVPYKVPAGVSLAARLK 391  
Qy 361 VSNMPOAVLPRAPNMLALQVMWCGGMEDIPCSRGHYRKVPYKVPAGVSLAARLK 420  
Db 361 VSNMPOAVLPRAPNMLALQVMWCGGMEDIPCSRGHYRKVPYKVPAGVSLAARLK 420  
Qy 392 RVAEVMDEYAEVYQRRPEYRHLASGDVAQKKLRSSLNCKSFKMTKIAMDLPRFY 451  
Db 392 RVAEVMDEYAEVYQRRPEYRHLASGDVAQKKLRSSLNCKSFKMTKIAMDLPRFY 451  
Qy 421 RVAEVMDEYAEVYQRRPEYRHLASGDVAQKKLRSSLNCKSFKMTKIAMDLPRFY 480  
Db 421 RVAEVMDEYAEVYQRRPEYRHLASGDVAQKKLRSSLNCKSFKMTKIAMDLPRFY 480  
Qy 452 PVSEPPAAWGEIRNVGTGLCADTYHGALGSPRLLEGCVRGGEAANNMOVFTFTWREDI 511  
Db 452 PVSEPPAAWGEIRNVGTGLCADTYHGALGSPRLLEGCVRGGEAANNMOVFTFTWREDI 511  
Qy 512 RPDGPQHTKKRCPAISHTSPTVTLYDCHSMKGNQLMKRYKDKTLHYHVSQSCMDSCSDH 571  
Db 512 RPDGPQHTKKRCPAISHTSPTVTLYDCHSMKGNQLMKRYKDKTLHYHVSQSCMDSCSDH 571  
Qy 572 RIFMNTCNPSLSLTQOMLFERTNSTVLEKEN 601  
Db 572 RIFMNTCNPSLSLTQOMLFERTNSTVLEKEN 601

Qy	338	GLEIMNGEQVYISIPFVNMVGGRMEDIPCSRVIHTRKVPYVVPVGVSLAARNIKRYAEVW	397
Qy	338	GLEIMNGEQVYISIPFVNMVGGRMEDIPCSRVIHTRKVPYVVPVGVSLAARNIKRYAEVW	397
Db	241	GLEIHWGSGQVYISIPFVNMVGGMEDIPCSRVIHTRKVPYVVPVGVSLAARNIKRYAEVW	300
Qy	398	MDEVAEYVYQRRPEPRHLISAGVAQAOKLIRSLINCKSPFKPMFTKIAMDLPKFYPVPEPPA	457
Db	301	MDEVAEYVYQRRPEPRHLISAGDVAVQKTLIRSLINCKSPFKPMFTKIAMDLPKFYPVPEPPA	366
Qy	458	AAMGEIRNVGTGLCADTGHGALGSPLEEGCVRGHGEAAMNNMQVFTFWREDIRPGDPQ	517
Db	361	AAMGEIRNVGTGLCADTGHGALGSPLEEGCVRGHGEAAMNNMQVFTFWREDIRPGDPQ	420
Qy	518	HTKKRCPFAISHTSPVTLTYDCHSMKGNOLMKYRKDKTLYHPVSGSCMDCSESDHRIEMNT	577
Db	421	HTKKRCPFAISHTSPVTLTYDCHSMKGNOLMKYRKDKTLYHPVSGSCMDCSESDHRIEMNT	480
Qy	578	CNPSSLTQOQMLFEHTNSTVLEKENFN 603	
Db	481	CNPSSLTQOQMLFEHTNSTVLEKENFN 506	
RESULT 6			
AC	AAAG79781	standard; protein; 506 AA.	
AC	AAAG79781;		
XX	AAAG79781;		
XX	16-APR-2003	(first entry)	
DT			
XX			
De		Carbohydrate-associated protein (CHOP) -3.	
XX			
KW		Human; carbohydrate associated polypeptides; CHOP; cancer;	
KW		carbohydrate metabolism; diabetes; anaemia; hypoglycaemia; obesity;	
KW		glycogen storage disease; neurological dysfunction; gene therapy;	
KW		cell proliferation; actinic keratosis; arteriosclerosis; inflammation;	
KW		atherosclerosis; cirrhosis; hepatitis; psoriasis; autoimmune disorder;	
KW		acquired immunodeficiency syndrome; AIDS; humanized; transgenic;	
KW		Addison's disease; allergy; asthma; contact dermatitis; bronchitis; gout;	
KW		Hashimoto's thyroiditis; infection; reproduction; endometriosis;	
KW		polycystic ovary syndrome; ovarian hyperstimulation syndrome;	
KW		Down syndrome; cystic fibrosis; sickle cell anemia; thalassemia;	
KW		myocarditis; cardiomyopathy; Alzheimer's disease; Parkinson's disease;	
KW		dementia; depression; epilepsy; Tourette's disorder; schizophrenia;	
KW		central nervous system; cerebral palsy; mood; anxiety; knockin.	
XX			
OS		Homo sapiens.	
XX			
XX			
FT	Key	Location/Qualifiers	
FT	Peptide	3..503	
FT		/label= Acetylgalactosaminyl transferase	
FT		/note= "Identified by BLAST_DOMO"	
FT	Peptide	8..493	
FT		/label= Acetylgalactosaminyl transferase	
FT		/note= "Identified by BLAST_DOMO"	
FT	Peptide	11..492	
FT		/label= Acetylgalactosaminyl transferase	
FT		/note= "Identified by BLAST_DOMO"	
FT	Peptide	51..236	
FT		/label= Glycosyl transferase sequence	
FT		/note= "Identified by HMMER_PPRM"	
FT	Peptide	202..362	
FT		/label= N-acetylgalactosaminyl transferase	
FT		/note= "Identified by BLAST_PRODOW"	
FT	Region	364..402	
FT		/label= QXW lectin repeat	
FT		/note= "Identified by HMMER_PPRM"	
FT	Region	417..445	
FT		/label= QXW lectin repeat	
FT		/note= "Identified by HMMER_PPRM"	
FT	Region	456..495	
FT		/label= QXW lectin repeat	
FT		/note= "Identified by HMMER_PPRM"	
XX			

PN WO200297060-A2.  
 XX 05-DEC-2002.  
 XX 22-MAY-2002; 2002WO-US018354.  
 PF 25-MAY-2001; 2001US-0293768P.  
 PR 01-AUG-2001; 2001US-0309548P.  
 PR 23-AUG-2001; 2001US-0314400P.  
 PR 19-OCT-2001; 2001US-0343706P.  
 PR 07-DEC-2001; 2001US-0337999P.  
 XX (INCYT-) INCYTE GENOMICS INC.  
 PA Swarnakar A, Gorvad AE, Hafalia AJN, Duggan BM, Emerling BM,  
 PI Iacon CH, Nguyen DB, Lee EA, Yue H, Forsythe IJ, Li JX,  
 PI Thangavelu K, Walla NK, Burford N, Mason PM, Lal PG, Lee S,  
 PI Becha SD, Tang YT;  
 XX WPI: 2003-140462/13.  
 DR N-PSDB; ABA00833.  
 XX Novel human carbohydrate associated polypeptide, useful in diagnosis,  
 PT treatment and prevention of carbohydrate metabolism, cell proliferative,  
 PT autoimmune/inflammatory, reproductive, and neurological disorders.  
 PS Claim 1; Page 126-27; 141pp; English.  
 XX The sequences given in AAG79779-88 represent human carbohydrate  
 CC associated polypeptides (CHOP). The CHOP polypeptides and the nucleotide  
 CC sequences encoding them, are useful for diagnosing, treating and  
 CC preventing cancer, carbohydrate metabolic disorders (e.g. diabetes,  
 CC anaemia, hypoglycaemia, obesity, glycogen storage disease, neurological  
 CC dysfunctions), cell proliferative disorders (e.g. actinic keratosis,  
 CC arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, psoriasis),  
 CC autoimmune/inflammatory disorders (such as acquired immunodeficiency  
 CC syndrome (AIDS), Addison's disease, allergies, asthma, contact  
 CC dermatitis, bronchitis, gout, Hashimoto's thyroiditis), viral, bacterial,  
 CC fungal, parasitic, protozoal and helminthic infections, reproductive  
 CC disorders (e.g. endometriosis, polycystic ovary syndrome, ovarian  
 CC hyperstimulation syndrome), genetic disorders (e.g. Down syndrome, cystic  
 CC fibrosis, sickle cell anaemia, thalassemia), cardiac disorders (e.g.  
 CC myocarditis, cardiomyopathy), neurological disorders (e.g. Alzheimer's  
 CC disease, Parkinson's disease, dementia, depression, epilepsy, Tourette's  
 CC disorder, schizophrenia), and developmental disorders of central nervous  
 CC system (e.g. cerebral palsy), mental disorders (e.g. mood, anxiety). CHOP  
 CC proteins are useful in a number of drug screening techniques, and to  
 CC analyse the proteome of a tissue or cell type. CHOP cDNA is useful for  
 CC creating "knockin" humanized animals or transgenic animals to model human  
 CC diseases, in somatic or germ-line gene therapy, to generate a transcript  
 CC image of a tissue or cell type, for detecting differences in the  
 CC chromosomal location due to translocation, inversion, etc., among normal,  
 CC carrier or affected individuals, and as hybridization probes for mapping  
 CC naturally occurring genomic sequences  
 XX Sequence 506 AA;  
 SQ  
 Query Match 84.5%; Score 2771; DB 6; Length 506;  
 Best Local Similarity 100.0%; Pred. No. 8.2e-274;  
 Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 TQAGDANRGAIPDEMYYKRIPIPELOKADPSDFPSPVWAGGLFVADRKMFNGLGYDP 337  
 DB 181 TQAGDANRGAIPDEMYYKRIPIPELOKADPSDFPSPVWAGGLFVADRKMFNGLGYDP 240  
 QY 338 GLEIWMGEQYEISPKWMCGRMBDIPCSRVGHIYRKYVYKYPAVAGSLARNLKRVAEVM 397  
 DB 241 GLEIWMGEQYEISPKWMCGRMBDIPCSRVGHIYRKYVYKYPAVAGSLARNLKRVAEVM 300  
 QY 398 MDEYAEYIYORRPEYRHSAGDVAVQKRLRSINLCKSFKMFMTKLIAMDLPKYPVPEPA 457  
 DB 301 MDEYAEYIYORRPEYRHSAGDVAVQKRLRSINLCKSFKMFMTKLIAMDLPKYPVPEPA 360  
 QY 458 AAMGEIRNVGTGLCADTRKHALGSPRLBECVGRGGAANNNOVFPTTRREDIRPDPO 517  
 DB 361 AAMGEIRNVGTGLCADTRKHALGSPRLBECVGRGGAANNNOVFPTTRREDIRPDPO 420  
 QY 518 HTKKECPDAISHSPVLYCHSKMGQMLKRYRDKTLVYPSGSCMDCSESDHRIEMNT 577  
 DB 421 HTKKECPDAISHSPVLYCHSKMGQMLKRYRDKTLVYPSGSCMDCSESDHRIEMNT 480  
 QY 578 CNPSSLTQOMLFERTNSTVLEKFNRN 603  
 DB 481 CNPSSLTQOMLFERTNSTVLEKFNRN 506  
 RESULT 7  
 ID AAU07771 standard; protein; 535 AA.  
 AC AAU07771;  
 XX 04-DEC-2001 (first entry)  
 DT Human novel transferase protein, NHP #14.  
 DE Human, transferase; breast cancer; prostatic cancer; immunogen;  
 XX gene therapy; antisense.  
 KM Homo sapiens.  
 OS WO200164903-A2.  
 PN 07-SEP-2001.  
 XX 28-FEB-2001; 2001WO-US006460.  
 PF 29-FEB-2000; 2000US-0185920P.  
 PR 02-MAR-2000; 2000US-0186558P.  
 PR 24-MAR-2000; 2000US-0191849P.  
 XX (LEXI-) LEXICON GENETICS INC.  
 PA Donoho G, Hillbun E, Turner CA, Friedrich G, Abuhin A;  
 PI Zambrowicz B, Sande AT, Walke DW, Wilganowski NL, Hu Y, Kleke JA,  
 PI Potter DG;  
 XX WPI: 2001-550185/61.  
 DR N-PSDB; AAS12618.  
 XX Novel nucleic acid sequences encoding novel human proteins useful for  
 PT diagnosis, drug screening, clinical trial monitoring and treatment of  
 PT diseases and disorders.  
 PS Claim 7; Page 45-46; 60pp; English.  
 XX The invention relates to isolated nucleic acids encoding novel human  
 CC transferase proteins (NHP). The nucleic acids and proteins are useful for  
 CC diagnosis, drug screening, clinical trial monitoring and treatment of  
 CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be  
 CC useful for augmenting the efficacy of chemotherapeutic agents used in  
 CC treatment of breast or prostate cancer. The nucleic acid is also useful  
 CC in NHP gene regulation, and as antisense primers in amplification  
 CC reactions of NHP gene sequences. NHPs are useful for producing



CC antibodies. The present sequence represents a novel human transferase  
 CC which has sequence similarity to N-acetyl-galactosaminyltransferase  
 XX  
 SQ Sequence 535 AA;

Query Match 83.8%; Score 2746.5; DB 4; Length 535;  
 Best Local Similarity 94.6%; Pred. No. 2.9e-271;  
 Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

QY 98 MTDNERDOAYRENGFNIVYSDKISLRSLPDIDHPNCNSKRYLETLPNTSIIIPFNEG 157  
 DB 1 MTDNERDOAYRENGFNIVYSDKISLRSLPDIDHPNCNSKRYLETLPNTSIIIPFNEG 60  
 QY 158 WSSILRTVHSLVNSPPELVAEIVLVNDFSDREHLKKPLBYMMLPFSVRLTRTKKREG 217  
 DB 61 WSSILRTVHSLVNSPPELVAEIVLVNDFSDREHLKKPLBYMMLPFSVRLTRTKKREG 120  
 QY 218 IRTTMLGASVATGVDITFLDSHCANVMMLPPLDRIARNRKIVCEMIDVIDHDDEPRYE 277  
 DB 121 IRTTMLGASVATGVDITFLDSHCANVMMLPPLDRIARNRKIVCEMIDVIDHDDEPRYE 180  
 QY 278 TQAGDANRGAAPDWMYTKRIPIPELOKADPSDFESPVAAGLFAVDKMFELGAYDP 337  
 DB 181 TQAGDANRGAAPDWMYTKRIPIPELOKADPSDFESPVAAGLFAVDKMFELGAYDP 240  
 QY 338 GLEIWGGEQYEISFK-----VNMCGRMEDIPCSRV 368  
 DB 241 GLEIWGGEQYEISFKGLHMLPRVLSNSMPOAVFLPRAPNMLALQVNMCGRMEDIPCSRV 300  
 QY 369 GHYRKVPKVPKPAVSLAENLKRAVEMWDEVAEYIYQRRPEYRHLISAGDVAVQKRLS 428  
 DB 301 GHYRKVPKVPKPAVSLAENLKRAVEMWDEVAEYIYQRRPEYRHLISAGDVAVQKRLS 360  
 QY 429 SLNCKSFKFMFTKIAMDLPKFYPPVEBPAAAMGSRINNVGTGLCADYKHALGSPRLREGC 488  
 DB 361 SLNCKSFKFMFTKIAMDLPKFYPPVEBPAAAMGSRINNVGTGLCADYKHALGSPRLREGC 420  
 QY 489 VRRGGEAANNMNOVFTTWRREDIRPGDPQHTKKCFPAISHTSPVTLVYDCHSMKGNOLMK 548  
 DB 421 VRRGGEAANNMNOVFTTWRREDIRPGDPQHTKKCFPAISHTSPVTLVYDCHSMKGNOLMK 480  
 QY 549 YRKDKTLVHPVSGSCMDCSDSDHRIFMNTCNPSLSLTQOMLFEHTNSTVLEKFNRN 603  
 DB 481 YRKDKTLVHPVSGSCMDCSDSDHRIFMNTCNPSLSLTQOMLFEHTNSTVLEKFNRN 535

RESULT 8  
 ID AAB94733 standard; protein; 339 AA.  
 AC AAB94733;  
 DT 26-JUN-2001 (first entry)  
 DB Human protein sequence SEQ ID NO:15766.  
 XX Human; primer; detection; diagnosis; antilease therapy; gene therapy.  
 XX Homo sapiens.  
 XX BP1074617-A2.  
 XX 07-FEB-2001.  
 XX 28-JUL-2000; 2000BP-00116126.  
 XX 29-JUL-1999; 99JB-00248036.  
 XX 27-AUG-1999; 99JB-00300253.  
 XX 11-JAN-2000; 2000JP-00118776.  
 XX 02-MAY-2000; 2000JP-00183767.  
 XX 09-JUN-2000; 2000JP-00241899.  
 XX (HELI-) HELIX RES INST.

XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.

PS Claim 8; SEQ ID NO 15766; 2537bp + Sequence Listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH16629 to AAH16632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention

XX SQ Sequence 339 AA;

Query Match 57.7%; Score 1893; DB 4; Length 339;  
 Best Local Similarity 99.7%; Pred. No. 3.1e-184;  
 Matches 338; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 265 MIDVIDHDDPRYETQGDANRGAAPDWMYTKRIPIPELOKADPSDFESPVAAGLFAV 324  
 DB 1 MIDVIDHDDPRYETQGDANRGAAPDWMYTKRIPIPELOKADPSDFESPVAAGLFAV 60  
 QY 325 DRKMFELAGYDPLGLIHWGGEQYEISFKVMMCGRMEDIPCSRGHYRKVPKVPAGV 384  
 DB 61 DRKMFELAGYDPLGLIHWGGEQYEISFKVMMCGRMEDIPCSRGHYRKVPKVPAGV 120  
 QY 385 SLARNLKRAVEMWDEVAEYIYQRRPEYRHLISAGDVAVQKRLSRLNCKSFKFMFTKIAM 444  
 DB 121 SLARNLKRAVEMWDEVAEYIYQRRPEYRHLISAGDVAVQKRLSRLNCKSFKFMFTKIAM 180  
 QY 445 DLKFPYPPVBPAAANGELRNVTGLCADYKHALGSPRLREGCVRGGAANNMNOVFT 504  
 DB 181 DLKFPYPPVBPAAANGELRNVTGLCADYKHALGSPRLREGCVRGGAANNMNOVFT 240  
 QY 505 FTWRREDIRPGDPQHTKKCFDAISHTSPVTLVYDCHSMKGNOLMKYRKDKTLVHPVSGSCM 564  
 DB 241 FTWRREDIRPGDPQHTKKCFDAISHTSPVTLVYDCHSMKGNOLMKYRKDKTLVHPVSGSCM 300  
 QY 565 DCSDSDHRIFMNTCNPSLSLTQOMLFEHTNSTVLEKFNRN 603  
 DB 301 DCSDSDHRIFMNTCNPSLSLTQOMLFEHTNSTVLEKFNRN 339

RESULT 9  
 ID AAU07776 standard; protein; 366 AA.  
 AC AAU07776;  
 XX AAU07776;

DT 04-DEC-2001 (first entry)  
 XX Human novel transferase protein, NHP #19.  
 XX Human; transferase; breast cancer; prostate cancer; immunogen;  
 KM gene therapy; anticense.  
 XX Homo sapiens.  
 OS  
 PN WO200164903-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 28-FEB-2001; 2001WO-US006460.  
 XX  
 PR 29-FEB-2000; 2000US-0185920P.  
 PR 02-MAR-2000; 2000US-0186558P.  
 PR 24-MAR-2000; 2000US-0191849P.  
 PA (LEXI-) LEXICON GENETICS INC.  
 XX  
 PI Donoho G, Hilbun E, Turner CA, Friedrich G, Abuln A;  
 PI Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kieke JA;  
 PI Potter DG;  
 XX WPI; 2001-550185/61.  
 DR N-PSDB; AAS12623.  
 XX  
 PT Novel nucleic acid sequences encoding novel human proteins useful for  
 PT diagnosis, drug screening, clinical trial monitoring and treatment of  
 PT diseases and disorders.  
 XX  
 PS Disclosure; Page 52-53; 60pp; English.  
 XX  
 CC The invention relates to isolated nucleic acids encoding novel human  
 CC transferase proteins (NHP). The nucleic acids and proteins are useful for  
 CC diagnosis, drug screening, clinical trial monitoring and treatment of  
 CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be  
 CC useful for augmenting the efficacy of chemotherapeutic agents used in  
 CC treatment of breast or prostate cancer. The nucleic acid is also useful  
 CC in NHP gene regulation, and as antisense primers in amplification  
 CC reactions of NHP gene sequences. NHPs are useful for producing  
 CC antibodies. The present sequence represents a novel human transferase  
 CC which has sequence similarity to N-acetyl-galactosaminyltransferase  
 CC  
 XX  
 SQ Sequence 366 AA;  
 Query Match 57.3%; Score 1877; DB 4; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-182;  
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRRREKRLQVALVLAALVLLPNVGLMALYREROPDGTGGSGAAVAPAGQSGSHSRQK 60  
 DB 1 MRRREKRLQVALVLAALVLLPNVGLMALYREROPDGTGGSGAAVAPAGQSGSHSRQK 60  
 QY 61 KTFPLGDOGKLMDHDKAIRDADQVNGEGRGPRYPTDAERYDQAVRENGFNIVYSDK 120  
 DB 61 KTFPLGDOGKLMDHDKAIRDADQVNGEGRGPRYPTDAERYDQAVRENGFNIVYSDK 120  
 QY 121 ISLRSRLPDIRHPNCNKRILETLPTNTSIIIPFNHGGSSILRTYHSLVNSPPPLVAEI 180  
 DB 121 ISLRSRLPDIRHPNCNKRILETLPTNTSIIIPFNHGGSSILRTYHSLVNSPPPLVAEI 180  
 QY 121 ISLRSRLPDIRHPNCNKRILETLPTNTSIIIPFNHGGSSILRTYHSLVNSPPPLVAEI 180  
 DB 121 ISLRSRLPDIRHPNCNKRILETLPTNTSIIIPFNHGGSSILRTYHSLVNSPPPLVAEI 180  
 QY 181 VLVDPSRRLKPLLEDYMALEPSVRLRTKKEGRLRTYMLGASVATGVTFFLDSHC 240  
 DB 181 VLVDPSRRLKPLLEDYMALEPSVRLRTKKEGRLRTYMLGASVATGVTFFLDSHC 240  
 QY 181 VLVDPSRRLKPLLEDYMALEPSVRLRTKKEGRLRTYMLGASVATGVTFFLDSHC 240  
 DB 181 VLVDPSRRLKPLLEDYMALEPSVRLRTKKEGRLRTYMLGASVATGVTFFLDSHC 240  
 QY 241 EAVNWMLEPLLDRILARNKRTIVCPMIDVIDHDFRYETQAGDAMKAGFDMWYTKRIP 300  
 DB 241 EAVNWMLEPLLDRILARNKRTIVCPMIDVIDHDFRYETQAGDAMKAGFDMWYTKRIP 300  
 QY 301 PELQKADSPDPFESPVMAGLFAVDRKMFELGSDPGLIHWGEQYEISERV 353  
 DB 301 PELQKADSPDPFESPVMAGLFAVDRKMFELGSDPGLIHWGEQYEISERV 353

DB 301 PELQKADSPDPFESPVMAGLFAVDRKMFELGSDPGLIHWGEQYEISERV 353  
 RESULT 10  
 ID AAG62600 standard; protein; 407 AA.  
 XX  
 AC AAG62600;  
 XX  
 DT 06-SEP-2001 (first entry)  
 XX  
 DE Human acetyl galactosyl transferase 45.  
 XX  
 KM Human; acetyl galactosyl transferase 45; GalNAc-T45; cancer; haemopathy;  
 KM HIV infection; immunological disease; inflammation; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200138545-A1.  
 XX  
 PD 31-MAY-2001.  
 XX  
 PR 20-NOV-2000; 2000WO-CN000473.  
 PR 24-NOV-1999; 99CN-00124100.  
 PA (BIOR-) BIOROAD GENE DEV LTD SHANGHAI.  
 XX  
 PI Mao Y, Xie Y;  
 XX WPI; 2001-355942/37.  
 DR N-PSDB; AAH45701.  
 XX  
 PT Human acetyl galactosyl transferase 45 and encoded polynucleotide, used  
 PT in diagnosis and treatment of malignant tumors, hemopathy, human  
 PT immunodeficiency virus infection, immunological diseases and  
 PT inflammation.  
 XX  
 PS Claim 1; Page 20-21; 33pp; Chinese.  
 XX  
 CC The present invention provides the protein and coding sequences of the  
 CC human acetyl galactosyl transferase 45 (GalNAc-T45). The sequences can be  
 CC used in the treatment of cancer, haemopathy, HIV infection, immunological  
 CC diseases and inflammation. The present sequence is the protein of the  
 CC invention  
 XX  
 SQ Sequence 407 AA;  
 Query Match 55.5%; Score 1818; DB 4; Length 407;  
 Best Local Similarity 78.5%; Pred. No. 2e-176;  
 Matches 317; Conservative 36; Mismatches 51; Indels 0; Gaps 0;  
 QY 200 MALPFSVRLIRTKRKGILRTYMLGASVATGVTFFLDSHC EAVNWMLEPLLDRILARNK 259  
 DB 1 MARFSKRLVIRTKRKGILRTYMLGASVATGVTFFLDSHC EAVNWMLEPLLDRILARNK 60  
 QY 260 TIVCPMIDVIDHDFRYETQAGDAMKAGFDMWYTKRIPPELQKADSPDPFESPVMAG 319  
 DB 61 TIVCPMIDVIDHDFRYETQAGDAMKAGFDMWYTKRIPPELQKADSPDPFESPVMAG 120  
 QY 320 GLFAVDRKMFELGSDPGLIHWGEQYEISFKVMCGGEMEDIPGRVGHITKTYPYK 379  
 DB 121 GLFAVDRKMFELGSDPGLIHWGEQYEISFKVMCGGEMEDIPGRVGHITKTYPYK 180  
 QY 380 VPAGVSLARNLKVAYVMDSEYATYVORREYHLSAGDVAVOKRLRSSLNCKSFKWM 439  
 DB 181 VPAGVSLARNLKVAYVMDSEYATYVORREYHLSAGDVAVOKRLRSSLNCKSFKWM 240  
 QY 440 TKIAMDLPKFPYPPVPAAGGEIRNYGTGICADTKHAGLSPLRLEGCVRGGEAAMNN 499  
 DB 241 AAAMVDPKTYPPVPPVPAAGGEIRNYGTGICADTKHAGLSPLRLEGCVRGGEAAMNN 300  
 QY 500 MOVFTFWREDIRGDDPQHTKKFCFDALSHTSPTVLTLDCHSMKGNLMTYRKDKTLVHPV 559



Db 301 EQFTFGMRDIRGEPDLHTRRKCFDAISHNSSPTLVLDCHGMKNQJMGARKORTLHPV 360  
 QY 560 SSGCMDCSESDHRIFMNTCNPSLITQOMLFPHNSTYLEKFNKN 603  
 Db 361 SNSCMDCNPAAKKIFMARCDPLSETQOMIFEHIMTVLEKFNH 404

## RESULT 11

ABG08187  
 ID ABG08187 standard; protein; 319 AA.

XX AEG08187;

XX 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #8178.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX MO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001MO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEO INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS72374.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensic, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

XX Claim 20; SEQ ID NO 38546; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridization probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensic, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 319 AA;

Query Match 42.8%; Score 1403.5; DB 4; Length 319;  
 Best Local Similarity 90.8%; Pred. No. 3.5e-134;  
 Matches 258; Conservative 4; Mismatches 19; Indels 3; Gaps 2;

QY 322 FAVDRKFWELGEGYDPEGLBIWGEQYELSPKVMCGGRMEDIPCSRIGHIYRKVPYKVP 381  
 Db 12 FAPCAGCWP-GAMDELEAETSSSLFLAMQVMCGGRMEDIPCSRIGHIYRKVPYKVP 70  
 QY 382 AGVSLAR--NLKRAVAEVMDEVAEYIYQRRPEYRHLSAGDVAVQKLRSSLNCKSPKFM 439  
 Db 71 AGVSLARSKYLKKAIVAMVDEVAEYIYQRRPEYRHLSAGDVAVQKLRSSLNCKSPKFM 130  
 QY 440 TKIAMDLPKFPYPPVEPPAAWGEIRNVGTGLCADTKHAGLSPRLREGYRGEAAWNN 499  
 Db 131 TKIAMDLPKFPYPPVEPPAAWGEIRNVGTGLCADTKHAGLSPRLREGYRGEAAWNN 190  
 QY 500 MOVPTFTMRDIRPDPQHKKCFDAISHTSPVLYDCHSMKNQJMGARKORTLHPV 559  
 Db 191 MOVPTFTMRDIRPDPQHKKCFDAISHTSPVLYDCHSMKNQJMGARKORTLHPV 250  
 QY 560 SSGCMDCSESDHRIFMNTCNPSLITQOMLFPHNSTYLEKFNKN 603  
 Db 251 SSGCMDCSESDHRIFMNTCNPSLITQOMLFPHNSTYLEKFNKN 294

## RESULT 12

ADE25865  
 ID ADE25865 standard; protein; 275 AA.

XX ADE25865;

XX 29-JAN-2004 (first entry)

XX Human N-acetyl]galactosaminyl transferase T10.

XX cytostatic; respiratory; antiaesthetic; antiinflammatory; endocrine; CNS;  
 KM immunosuppressive; antiarthritic; antibacterial; vulnery;  
 KM GalNAc-transferase lectin domain inhibitor;  
 KM N-acetyl]galactosaminyl transferase; GalNAc-transferase; GalNAc-T4;  
 KM GalNAc-T7; GalNAc-T2; GalNAc-T3; GalNAc-transferase; tumour; cancer;  
 KM lung disease; mucous accumulation; asthma; chronic bronchitis;  
 KM smoker's lung; cystic fibrosis; exocrine gland disease;  
 KM mucin secretion decrease; Sjogren's syndrome; dry mouth;  
 KM selection-mediated leukocyte trafficking; autoimmunity; arthritis;  
 KM leukaemia; lymphoma; immunosuppression; sepsis; wound healing;  
 KM inflammation; mucin secretion inhibition; hypersecretion;  
 KM mucin accumulation; chronic obstructive respiratory pulmonary disease;  
 KM asthma; lectin domain inhibition; GalNAc-glycopeptide;  
 KM Galbeta1-3GalNAc-glycopeptide; UDP-Gal donor; glycosyltransferase;  
 KM core 1 beta1,3-galactosyltransferase; alpha2,6-sialyltransferase;  
 KM N-acetyl]galactosaminyl transferase 4Ap24 glycopeptide; O-glycosylation;  
 KM breast cancer; human; GalNAcT10.

XX Homo sapiens.

XX US2003186850-A1.

XX 02-OCT-2003.

XX 12-NOV-2002; 2002US-00292896.

XX 11-MAY-2000; 2000US-0203311P.

XX 10-MAY-2001; 2001MO-DK000328.

XX 08-NOV-2002; 2002US-0425204P.

XX (GLYC-) GLYCOZYX APS.

XX Clausen H, Bennett EP, Hasaan H, Reis CA;

XX WPI; 2003-831259/77.

XX Modulating N-acetyl]galactosaminyltransferase functions comprises  
 PT administration of an N-acetyl]galactosaminyltransferase lectin domain  
 PT inhibitor.

PS Disclosure; Page 11; 65pp; English.

XX The invention describes a method of modulating polypeptide N-  
CC acetylglucosaminyl (GALNAc)-transferase functions comprising  
CC administration of a GALNAc-transferase lectin domain inhibitor (I). (I)  
CC is used for modulating functions of GALNAc-transferase (e.g. GALNAc-T4,  
CC GALNAc-T7, GALNAc-T2, and GALNAc-T3); for inhibiting at least one lectin  
CC domain of the GALNAc-transferase and modulating the function mediated by  
CC the domain for preparing a medicament for treating tumours and cancers,  
CC lung diseases associated with mucous accumulation (e.g. asthma, chronic  
CC bronchitis, smoker's lung, and cystic fibrosis), diseases of exocrine  
CC glands associated with increased or decreased mucin secretion (e.g.  
CC Sjogren's syndrome and dry mouth), disorders associated with  
CC dysregulation of selectin-mediated leukocyte trafficking (e.g.  
CC autoimmunity, arthritis, leukemias, lymphomas, immunosuppression,  
CC sepsis, wound healing, acute and chronic inflammation); inhibiting mucin  
CC secretion and inhibiting hypersecretion and accumulation of mucin in the  
CC lungs of mammals (e.g. humans) suffering from chronic obstructive  
CC respiratory pulmonary diseases, asthma, and cystic fibrosis. (I) is  
CC effective in inhibiting at least one lectin domain of the GALNAc-  
CC transferase and modulating the inhibited function mediated by the domain  
CC (preferably GALNAc-glycopeptide, galbetail-3GALNAc-glycopeptide, or  
CC peptide specificity of the transferase, or UDP-gal donor substrate  
CC specificity). (I) selectively inhibits at least one member of the GALNAc-  
CC transferase family without inhibiting other glycosyltransferases  
CC (preferably core 1 beta1,3-galactosyltransferase, alpha2,6-  
CC sialyltransferase, or glycosyltransferase functioning in the O-  
CC glycosylation pathway). This is the amino acid sequence of human N-  
CC acetylglucosaminyl transferase T10.

XX Sequence 275 AA;

Query Match 42.6%; Score 1396; DB 7; Length 275;  
Best Local Similarity 93.7%; Pred. No. 1.6e-133;  
Matches 253; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 336 DPGLEIMGGEYYSFKYWMCGRMEDIPCSRVGHIYRKYVPYKVPAGVSLA--RNLKRV 393  
DB 6 DGELEETSSIFLLAMQVMCGRMEDIPCSRVGHIYRKYVPYKVPAGVSLARVTTLRV 65  
QY 394 AEVMMDEVAEYIYORRPREYRLHLSAGDVAVOKKRLSLNCKSPKWMPTIAMDLPKFPYPV 453  
DB 66 AEVMMDEVAEYIYORRPREYRLHLSAGDVAVOKKRLSLNCKSPKWMPTIAMDLPKFPYPV 125  
QY 454 EPPAAAMGEIRNVGTGLCADTKHGALGSPRLRLGCVRGGEAAMNNQVFTFTWRREDIRP 513  
DB 126 EPPAAAMGEIRNVGTGLCADTKHGALGSPRLRLGCVRGGEAAMNNQVFTFTWRREDIRP 185  
QY 514 GDPQHTKKFCFDALSHTSPTVLYDCHSMKGNQLMKTRKDKTLVHPVSGSCMDCESDHRI 573  
DB 186 GDPQHTKKFCFDALSHTSPTVLYDCHSMKGNQLMKTRKDKTLVHPVSGSCMDCESDHRI 245  
QY 574 FMNTCNPSLTQOWLFETHNSTVLEKFN 603  
DB 246 FMNTCNPSLTQOWLFETHNSTVLEKFN 275

RESULT 13

ADQ17613

ID ADQ17613 standard; protein, 276 AA.

XX ADQ17613;

DT 26-AUG-2004 (first entry)

DB Human soft tissue sarcoma-upregulated protein - SEQ ID 430.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.

OS Homo sapiens.

KM WO2004048938-A2.

XX 10-JUN-2004.

PD

XX 26-NOV-2003; 2003WO-US038193.  
XX 26-NOV-2002; 2002US-0429739P.  
XX (PROT-) PROTEIN DESIGN LABS INC.  
XX Aziz N, Gineburg WM, Zlotnik A;  
XX WPI; 2004-441208/41.

PT Early detection of soft tissue sarcoma comprises determining expression  
PT of a gene in a first soft tissue sample and a normal soft tissue sample  
PT and comparing the gene expression, also useful in treating soft tissue  
PT sarcoma.

PS Example 2; SEQ ID NO 430; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma  
CC which comprises obtaining a first soft tissue sample from an individual  
CC and a normal soft tissue sample from the same or different individual,  
CC determining the expression of a gene in both samples and comparing the  
CC expression of the gene in both soft tissue samples, where a higher level  
CC of protein expression in the first soft tissue sample indicates the  
CC presence of soft tissue sarcoma. The method of the invention has  
CC cytostatic applications and may be useful for detecting soft tissue  
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
CC acid sequences may be useful in diagnostic and screening applications.  
CC The current sequence is that of a human soft tissue sarcoma-upregulated  
CC protein of the invention. The current sequence is not shown within the  
CC specification per se but was submitted in CD format by the inventor.

SQ Sequence 276 AA;

Query Match 42.5%; Score 1392; DB 8; Length 276;  
Best Local Similarity 97.3%; Pred. No. 4.2e-133;  
Matches 250; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 349 ISFKVMCGRMEDIPCSRVGHIYRKYVPYKVPAGVSLA--RNLKRVAEVMMDEVAEYIY 406  
DB 20 LAMQVMCGRMEDIPCSRVGHIYRKYVPYKVPAGVSLARVTTLRVAEVMMDEVAEYIY 79  
QY 407 QRRPREYRLHLSAGDVAVOKKRLSLNCKSPKWMPTIAMDLPKFPYVPEPPAAAMGEIRNV 466  
DB 80 QRRPREYRLHLSAGDVAVOKKRLSLNCKSPKWMPTIAMDLPKFPYVPEPPAAAMGEIRNV 139  
QY 467 GTGACADTKHGALGSPRLRLGCVRGGEAAMNNQVFTFTWRREDIRPQDPTKKFCFDA 526  
DB 140 GTGACADTKHGALGSPRLRLGCVRGGEAAMNNQVFTFTWRREDIRPQDPTKKFCFDA 199  
QY 527 ISHTSPVTLVYDCHSMKGNQLMKTRKDKTLVHPVSGSCMDCESDHRIFMNTCNPSLTQ 586  
DB 200 ISHTSPVTLVYDCHSMKGNQLMKTRKDKTLVHPVSGSCMDCESDHRIFMNTCNPSLTQ 259  
QY 587 WLFETHNSTVLEKFN 603  
DB 260 WLFETHNSTVLEKFN 276

RESULT 14

ABG08189

ID ABG08189 standard; protein, 335 AA.

XX ABG08189;

DT 13-FEB-2002 (first entry)

DB Novel human diagnostic protein #8180.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS

XX WO200175067-A2.  
 PN 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US008631.  
 XX 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX (HXYE-) HXYEQ INC.  
 PA Dmanac RT, Liu C, Tang YT;  
 PI WPI, 2001-639362/73.  
 XX N-PSDB; AAS172376.  
 DR  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 CC Claim 20; SEQ ID NO 38548; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC tcp.wipo.int/pub/published\_pct\_sequences  
 CC  
 XX Sequence 335 AA;  
 SQ  
 Query Match 42.3%; Score 1385.5; DB 4; Length 335;  
 Best Local Similarity 87.5%; Pred. No. 2.7e-132;  
 Matches 253; Conservative 0; Mismatches 1; Indels 35; Gaps 1;

QY 350 SFAKVMCGHMEIDPCSRVGHVYKYPVAGVSLARNLKRAVAVMDEAYEYIYORR 409  
 Db 47 SHKVMCGHMEIDPCSRVGHVYKYPVAGVSLARNLKRAVAVMDEAYEYIYORR 106  
 QY 410 PEYRHLSAGDVAVOKLRSINCKSPKFMPTKIAMDLPKFYPPVPPAAANG----- 462  
 Db 107 PEYRHLSAGDVAVOKLRSINCKSPKFMPTKIAMDLPKFYPPVPPAAANG----- 166  
 QY 463 -----IRVVGICADTGHGALSGPLRLEGVVRGGE 494  
 Db 167 YSKGFMSFSRRFQGMVBPAPAPSPGMCIRNVGTGLCADTGHGALSGPLRLEGVVRGGE 226  
 QY 495 AAMNMVFTTWRBEDIKPGDPQHTKKFCFPAISHTSPVTLTDCHSMKGNOLATYRXDKT 554  
 Db 227 AAMNMVFTTWRBEDIKPGDPQHTKKFCFPAISHTSPVTLTDCHSMKGNOLATYRXDKT 286  
 QY 555 LYHVPVSGSCWDCCSSDHRIFMNTCNPSLSLTOQMLFEHTNSTVLEKFNRN 603  
 Db 287 LYHVPVSGSCWDCCSSDHRIFMNTCNPSLSLTOQMLFEHTNSTVLEKFNRN 335

RESULT 15  
 ID AAU07770 standard; protein; 269 AA.  
 XX AAU07770;  
 AC 04-DEC-2001 (first entry)  
 DT 04-DEC-2001 (first entry)  
 XX Human novel transferase protein, NHP #13.  
 DE Human; transferase; breast cancer; prostate cancer; immunogen;  
 KW gene therapy; antisense.  
 XX Homo sapiens.  
 OS  
 PN WO200164903-A2.  
 PD 07-SEP-2001.  
 XX 28-FEB-2001; 2001WO-US006460.  
 PR 29-FEB-2000; 2000US-0185920P.  
 PR 02-MAR-2000; 2000US-0186558P.  
 PR 24-MAR-2000; 2000US-0191849P.  
 XX (TEXT-) LEXICON GENETICS INC.  
 PA Donoho G, Hilborn B, Turner CA, Friedrich G, Abuin A;  
 PI Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kieke JA;  
 PI Potter DG;  
 XX WPI, 2001-550185/61.  
 DR N-PSDB; AAS12617.  
 DR  
 XX Novel nucleic acid sequences encoding novel human proteins useful for  
 PT diagnosis, drug screening, clinical trial monitoring and treatment of  
 PT diseases and disorders.  
 XX  
 XX Disclosure; Page 44-45; 60pp; English.  
 PS  
 XX The invention relates to isolated nucleic acids encoding novel human  
 CC transferase proteins (NHP). The nucleic acids and proteins are useful for  
 CC diagnosis, drug screening, clinical trial monitoring and treatment of  
 CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be  
 CC useful for augmenting the efficacy of chemotherapeutic agents used in  
 CC treatment of breast or prostate cancer. The nucleic acid is also useful  
 CC in NHP gene regulation, and as antisense primers in amplification  
 CC reactions of NHP gene sequences. NHPs are useful for producing  
 CC antibodies. The present sequence represents a novel human transferase  
 CC which has sequence similarity to N-acetyl-galactosaminyltransferase  
 CC  
 XX Sequence 269 AA;  
 SQ  
 Query Match 41.8%; Score 1370; DB 4; Length 269;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-131;  
 Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDABRVDAQYRBNFNIVYSDKISLNRSLPDIRHPCNCKRYLETLPNTSIIIPFNEG 157  
 Db 1 MTDABRVDAQYRBNFNIVYSDKISLNRSLPDIRHPCNCKRYLETLPNTSIIIPFNEG 60  
 QY 158 WSSLIRTVSHVILNRSPPELVAEIVLVDDPSDRHLKKPLDYVALPFSVILRTKREG 217  
 Db 61 WSSLIRTVSHVILNRSPPELVAEIVLVDDPSDRHLKKPLDYVALPFSVILRTKREG 120  
 QY 218 IRTBMLGASVATGVITFLSHCEANVMPLPLDIRIARRKTYICPMIDVIOHDDRYE 277  
 Db 121 IRTBMLGASVATGVITFLSHCEANVMPLPLDIRIARRKTYICPMIDVIOHDDRYE 180  
 QY 278 TOAGDAMRGAFDWEMYYKRIPIPELOKADPSDFSPVVAAGLFAVDRKMFPELGYPD 337  
 Db 181 TOAGDAMRGAFDWEMYYKRIPIPELOKADPSDFSPVVAAGLFAVDRKMFPELGYPD 240

QY 338 GLBIWGEQYEIFKV 353  
 Db 241 GLBIWGEQYEIFKV 256

RESULT 16  
 ID ABB66873 standard; protein; 666 AA.  
 XX ABB66873;

AC ABB66873;  
 XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 27411.

XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

XX Drosophila melanogaster.

XX MO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL10976.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

PS Disclosure; SEQ ID NO 27411; 21bp + Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_sequences

XX Sequence 666 AA;

Query Match 41.5%; Score 1360; DB 4; Length 666;  
 Best Local Similarity 49.0%; Pred. No. 3.3e-129;  
 Matches 277; Conservative 76; Mismatches 160; Indels 52; Gaps 15;

QY 67 DQGLKMDHDKAARRDAQVNGEGRPYMTDAERVDQVR--ENGFNITYSDKISL 123

Db 121 DASVAKDMDHDTYFMEKDKARVQLGEGKASTLDDESGQRLDKRMSLENGFPALSDSISV 180

QY 124 NRSI.PDIIRHPNCSKRYLETLPNTSIIIPFNEGSSILRTVSHVSLNPSPELVAEIVLY 183

Db 181 NRSVPDIDHPLCRKKEVYAKLPYVSVIIIFNEVYSLMRSHSLINSPELMKEITLV 240

QY 184 DDFSDREHLKKPLBDYMA-LFPYVAILRTKRGGLIRTRMGASVATGVDVITFLDSHCEA 242

Db 241 DDHSDREVLGKLELYIAEHFKWAVVLRPRGTGIGARAGARVATAEVLIFLDSHVEA 300

QY 243 NVNMLPRLLDRIANRRKTIYVCPMIDVIVHDDFRYTGAGDAMRGAFEMWTYKRIPIPE 302

Db 301 NVNMLPRLLEPIALNKRITAVCPFIDVITDHTNFHYRAQ-DEGARGAFMEFFYKRLPLLPE 359  
 QY 303 LQKADPSDPSPSPVMAAGLFAVDKRWFWELGYPDGLIWMGEQYEISPKVMCGRMED 362  
 Db 360 DLK-HPADPFKSPFMAGLFAISRFWELGIDEGDIDMGREYELSFKIMWGGMND 418  
 QY 363 IPCSRGVHYR-----KRYPYKVPAGVSLARNLKRVAEVMDEYAEIYQRRPE-YHLS 416  
 Db 419 APCSRIGHIYRGRPHQPSDRK---GDYLHKNYKRVAEVMDEYKNLYSHGDLESVD 475  
 QY 417 AGDVAVQKLRSSLNCSFPMFKTIAMDLPKPYRPVPPAAMGELRNVT-GLCADT- 474  
 Db 476 PGDLTEQALRTKLNCSEFFWEEVAFDLMKTYPPVDPSPYAMGALQNGNONLCLDTL 535  
 QY 475 ---KGALG-----SPLRLEGCVRGREGAAMNNQVFTFTRREDIRPGDPQTKKF 522  
 Db 536 GRKGNMGMVACADNKTQR-----TQFWELSMKDLR---LRKKE 576  
 QY 523 CFDA-ISHTSPVTLYDCHSMKGNQLMKY-RKDKTYHPVSG-SCMDCSSDRIENMTC 578  
 Db 577 CLDVQWDANAPVWLMDCHSGGQGVYDYDRHKQLRGTEGRCLTLPPSQEVVANKC 636  
 QY 579 NPSSLTOQMLFEHTNSVLEKFNKN 603  
 Db 637 DTNRFQQMNFGSFNKTDLDNYSOD 661

RESULT 17  
 ABB58867  
 ID ABB58867 standard; protein; 666 AA.

XX ABB58867;

AC 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 3393.

XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

XX Drosophila melanogaster.

XX MO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL02970.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

PS Disclosure; SEQ ID NO 3393; 21bp + Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly

CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
XX  
SQ Sequence 666 AA;

**SQ** Sequence 666 AA;

Query Match	41.5%	Score 1360;	DB 4;	Length 666;
Best Local Similarly	49.0%	Pred. No. 3.3e-129;		
Matches 277; Conservative	76;	Mismatches 160;	Indels 52;	Gaps 15

[illegible]

RESULT	18
AAU07774	
ID	AAU07774 standard; protein; 321 AA.
XX	
AC	AAU07774;
XX	
DT	04-DEC-2001 (first entry)
XX	
DE	Human novel transferrase protein, NHP #17.
XX	
KM	Human; transferrase; breast cancer; prostate cancer; immunogen;
XX	
OS	gene therapy; antisense.
XX	
OS	Homo sapiens.
XX	
PN	W0200164903-A2.
XX	
PD	07-SEP-2001.
XX	
PF	28-FEB-2001; 2001WO-US0006460.
XX	
PR	29-FEB-2000; 2000US-0185920P.
XX	
PR	02-MAR-2000; 2000US-0186558P.
XX	
PR	24-MAR-2000; 2000US-0191849P.
XX	

XX  
PA (LEXI-) LEXICON GENETICS INC.

PI Donoho G, Hilbun B, Turner CA, Friedrich G, Abulin A;  
PI Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kieke JA;  
PI Potter DG;

DR WPI; 2001-550185/61.  
DR N-PSDB; AAS12621.

PT Novel nucleic acid sequences encoding novel human proteins useful for  
PT diagnosis, drug screening, clinical trial monitoring and treatment of  
PT diseases and disorders.

PS Disclosure; Page 50-51; 60pp; English.

CC The invention relates to isolated nucleic acids encoding novel human  
CC transferase proteins (NHP). The nucleic acids and proteins are useful for  
CC diagnosis, drug screening, clinical trial monitoring and treatment of  
CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be  
CC useful for augmenting the efficacy of chemotherapeutic agents used in  
CC treatment of breast or prostate cancer. The nucleic acid is also useful  
CC in NHP gene regulation, and as antisense primers in amplification  
CC reactions of NHP gene sequences. NHPs are useful for producing  
CC antibodies. The present sequence represents a novel human transferase  
CC which has sequence similarity to N-acetyl-galactosaminyltransferase

**SQ Sequence 321 AA;**

Query Match	40.0%	Score 1311	DB 4	Length 321
Best Local Similarly	99.2%	Pred. No. 1e-124		
Matches 251, Conservative	1	Mismatches	1	Indels 0
				Gaps 0

QY	1	MRKKRRLIQAVLVLAALVLLPNVGLMALYEROPDGTFGSSGAAVVAAPAGGSSHSRQK	60
Db	1	MRKKRRLIQAVLVLAALVLLPNVGLMALYEROPDGTFGSSGAAVVAAPAGGSSHSRQK	60
QY	61	KTFFLGDSQKLDKMDHDKAIRDAORVNGGEOGRPYPMTDARVDAQYRENGFNIVYSDK	120
Db	61	KTFFLGDSQKLDKMDHDKAIRDAORVNGGEOGRPYPMTDARVDAQYRENGFNIVYSDK	120
QY	121	ISLRSLPDIRHPNCSKRYLETLPNTSIIIPFNHGWSSLLRTYHSLVNRSPPELVAEI	180
Db	121	ISLRSLPDIRHPNCSKRYLETLPNTSIIIPFNHGWSSLLRTYHSLVNRSPPELVAEI	180
QY	181	VLVDDFSDBRHKKPELDEYMAFLPESVRLIRTKKREGILIRTMIGASVATGDTVITFLDSHC	240
Db	181	VLVDDFSDBRHKKPELDEYMAFLPESVRLIRTKKREGILIRTMIGASVATGDTVITFLDSHC	240
QY	241	EANVMVLPPLDRLDR	253
Db	241	EANVMVLPPLDRLDR	253

XX	RESULT 19
XX	ABB59363
ID	ABB59363 standard; protein; 1252 AA.
XX	
AC	ABB59363;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 4881.
XX	
KX	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	

PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-655860/75.  
DR N-PSDB; ABL03486.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 4881; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-  
CC AB572072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1252 AA;

Query Match 36.2%; Score 1186; DB 4; Length 1252;  
Best Local Similarity 41.2%; Pred. No. 5.8e-111;  
Matches 266; Conservative 106; Mismatches 216; Indels 58; Gaps 21;

QY 2 RRKERRLLQAVAVLVAALVLLPNVGLMALYR-----ERQPDG---TPGSGAAVA 48  
DB 613 KRYVKRLLRKVVLLVIVTVSLVTVVERRMKVAELTQQLDNPNDPITPFFRAANIH 672  
QY 49 PA--AGGQSHSRCKPTFLGQOQL-----KDMDKAIRDADRVNGEGGR 94  
DB 673 PTRAPRPFPQDMVSDIPRSDIKQGFRLPEPKGERDMDHVAEMEDRRSGEGEGV 732  
QY 95 PYPW--TDAERV--QAYRENGFNIVSDKISLNRSLPDIRPNCSSKRYLTLPNTSII 151  
DB 733 AVKLENDPEKLEKEHEMNGFNGLISRISVNSVPLRLKACTRTYAKLPRIISITF 792  
QY 152 PFHNEGSSLLRTVHSLVNSRPPELVAEIVLVDFSDREHLKKPLEDMAL-PPS-VRI 209  
DB 793 IFPNEHFWTLRSIYSVINRTPPELLKQIVLVDDGSEWDVKQPLDDYVQGHFPLVTV 852  
QY 210 RTKRGEGILRTRMGASVATGDVITFLDSHCEBANNVLPRLDRIARAKRTIVCPMDIV 269  
DB 853 RNPBROGILIGRIAGAKAVAGQVWFSDSHLVNNWMLPPIEPIAINPKISTGCMVDTI 912  
QY 270 DHDQFRYETGADMARGAFFDMEMVYKRIPIPELOKADPPSPFESPVNAGGLFAVDKMF 329  
DB 913 SHEPFSYSGNKDARGGDFMKMLYKQLPVLPB-DALDKMFRYSPVMGSLFALINDPF 971  
QY 330 WELGQYDGLIWIWGEQYEIFKVMWCGRMEDIPCSFVGHYR-KYVPYKVPAGVS-LA 387  
DB 972 WDLGGYDQDLIWIWGEQYELSPKIMWCGMILLDVCSVAHIFRPMKPRGNRSHNIFA 1031  
QY 388 RNLRKVAVMVDEVAEYIYQRPE-YRHLSDGDAVAVQKGLASSLNCSEFKKPMFTIANDL 446  
DB 1032 KHHKRVAVVMVDEYKQYVYKBDPKTYDMLDAGDLTRQGVBERLKCKSEFHHFMEVADF 1091  
QY 447 PKFYPPVPPPAAMEIRNVGTG-CADTKGALGSLRLSGCVAGREAAAMNNQVPTF 505  
DB 1092 LVKFPVPPBPSTAAITGVANPVYCLDNMGKSTEAIVGMSCADNRTHPOEN-QFWEL 1149  
QY 506 TWRREDIRGDDPOHTKKF---CFDAISHTSP---VTLYDCHSMKGNQULMKY-RKDKTIYH 557

DB 1150 SIFRDLE-----MKGDFSVCLDV--HEGPNATVNMWMSCHSQGQGFMYRQTORLVH 1201  
QY 558 PVSQS-CMD--CSRSDHRIFMATCNPSLSLTQQLFETNSTVLEKEF 600  
DB 1202 GENMKRCLGEGFVENGIAKVYVANGCENDGRORNEFGFVHTMTDITF 1247

RESULT 20  
ABG08190  
ID ABG08190 standard; protein; 499 AA.

AC ABG08190;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #8181.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN W0200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

XX N-PSDB; AAS72377.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PS Claim 20; SEQ ID NO 38549; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 499 AA;

Query Match 35.2%; Score 1155; DB 4; Length 499;  
Best Local Similarity 51.0%; Pred. No. 2e-108;  
Matches 251; Conservative 12; Mismatches 31; Indels 198; Gaps 10;

QY 175 ELVAEIVLVDPSREHLKKPLEDYMLPSVRILRTKKRGLRTBMLGASVATGVIT 234  
 DB 21 EMQGEITKLKDGSGHEHLKKPLEDYMLPSVRILRTKKRGLRTBMLGASVATGVIT 80  
 QY 235 FLDSHCANVNMPLPLDLRIARNKRTIVCPMIDVIDHDFRYETQADAMKGFDMEMYY 294  
 DB 81 FLDSHCANVNMPLPLDLRIARNKRTIVCPMIDVIDHDFRYETQADAMKGFDMEMYY 140  
 QY 295 KRPIPELOKADPSDPPE-----SPMAGGLF 322  
 DB 141 KRPIPELOKADPSDPPEQVLVDLGDADSOQHDVSHLARCQADFLFPMSPVMAAGLGF 200  
 QY 323 AVRRKFMWELGMY-----DP----- 337  
 DB 201 AVRRKFMWELGMYFSGHDPQPKESLFPPSSCTSPQVESHMPAGMPPIPEOYCNMSPA 260  
 QY 338 -----GLEIMWG-----EYVE-- 348  
 DB 261 VTCPVTAGRVCLGWMGRKXKMGTKDILGARWKSXKPPXPTNQDVKNPHDLIEGFQK 320  
 QY 349 ---ISPRVNM-----CGRMEDIPC 365  
 DB 321 EHCFTESAMVIGCPONGSMWLOSHREBNHGIEMTVISAVSCGLHGVCGGMEDIPC 380  
 QY 366 SRVGHIRKTVVYKVPAGVSLARLKRVAEVMWDEVAEYIYQRRPEYRHLISAGDVAVQKK 425  
 DB 381 S-----RNLKRVAEVMWDEVAEYIYQRRPEYRHLISAGDVAVQKK 419  
 QY 426 LRSSLNCKSFKFWMTKIAMDLPKFYPRVEPPAAAMGEIRNVGTGLCADTKHAGLSPLRL 485  
 DB 420 LRSSLNCKSFKFWMTKIAMDLPKFYPRVEPPAAAMGEIRNVGTGLCADTKHAGLSPLRL 476  
 QY 486 EGCY-----RGR 492  
 DB 477 GSCTHHWGRGR 488  
 RESULT 21  
 ID ADP65304 standard; protein, 559 AA.  
 XX ADP65304;  
 XX 12-AUG-2004 (first entry)  
 DE Human polypeptide N-acetylgalactosaminyltransferase 1.  
 XX autoimmunity disease; arthritis; gene expression analysis;  
 KM rheumatoid arthritis; collagen-induced; immunosuppressive; antirheumatic;  
 KM antirheumatic; osteoarthritis; gout; antiinflammatory; dermatological;  
 KM immunomodulatory; lupus; ankylosing spondylitis; fibrositis;  
 KM fibromyalgia; osteoarthritis; gout; juvenile rheumatoid arthritis;  
 KM immune; human.  
 OS Homo sapiens.  
 PN W02003072827-A1.  
 XX 04-SEP-2003.  
 PD 31-OCT-2002; 2002WO-US035433.  
 PF 31-OCT-2001; 2001US-0336220P.  
 PR (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.  
 PA Hirsch R. Thorton SL;  
 XX WPI, 2003-712740/67.  
 DR GENBANK; NP\_065207.  
 XX PT  
 PT profiles and microarray technology, useful for diagnosing and treating

PT rheumatoid arthritis, lupus, fibrositis, osteoarthritis, fibromyalgia and  
 PT gout.  
 XX Disclosure; Page; 56pp; English.  
 XX The invention relates to a novel method for diagnosing and analysing  
 CC autoimmune disease or arthritides. The method comprises obtaining a  
 CC patient sample containing mRNA, analysing gene expression using the mRNA  
 CC that results in a gene expression signature or analyse the autoimmune disease  
 CC or arthritides in the patient, where gene expression of at least 60% of  
 CC the genes correlates with that of the gene signature. The invention  
 CC further comprises: a treatment of rheumatoid arthritis; identification  
 CC of genes for targeting in the treatment of rheumatoid arthritis in a mammal  
 CC other than a mouse; diagnosis of rheumatoid arthritis in a mammal; an  
 CC array or gene chip, specific for rheumatoid arthritis; diagnosis or  
 CC analysis of autoimmune disease or rheumatoid arthritis; screening the  
 CC efficacy of a candidate drug in vitro for the treatment of collagen-  
 CC induced arthritis; and reducing the symptoms associated with collagen-  
 CC induced arthritis. The compositions of the invention have the following  
 CC activities: immunosuppressive, antirheumatic, antirheumatic, osteopathic,  
 CC antihistamine, antiinflammatory, dermatological, and immunomodulatory. The  
 CC methods and compositions of the present invention are useful for  
 CC diagnosing and treating autoimmune disease or arthritides, such as  
 CC rheumatoid arthritis, lupus, ankylosing spondylitis, fibrositis,  
 CC fibromyalgia, osteoarthritis, gout, juvenile rheumatoid arthritis, and an  
 CC immune disease caused by an infectious agent. This sequence represents a  
 CC protein sequence relating to the genes used in the analysis and treatment  
 CC of autoimmune diseases or arthritides. Note: This sequence is not shown  
 CC in the specification. It has been supplied in an electronic format from  
 CC WIPO.  
 XX  
 SQ Sequence 559 AA;  
 Query Match 34.3%; Score 1125; DB 7; Length 559;  
 Best Local Similarity 44.0%; Pred. No. 2,8e-105;  
 Matches 232; Conservative 87; Mismatches 170; Indels 38; Gaps 16;  
 QY 88 GNGEQRP--YPMTEARVDQAVRENGFNIVSDKISLNSLPDIRHPCNSKRYLETLP 145  
 DB 57 GPGEMGRKVPVIRKDEQKMKEMFKINQFNLMASEMILANSLDPVRLGCKTKVPPDNL 116  
 QY 146 NTSIIIPHNHSGSSILRTVSHVNSPPELVAEIVLVDPSREHLKKLEDMV--ALFP 204  
 DB 117 TTSVVIVFHNHSAWSTLRTVSHVNSPPELVAEIVLVDPSREHLKKLEDMV--ALFP 176  
 QY 205 SVRLRTKKRGLIRFTMLGASVATGVITFLDSHCANVNMPLPLDLRIARNKRTIVCP 264  
 DB 177 PAVIRHNEQSGILIRKARLKAASKGVITFLDHCCTGVWLEPLARIKHRRRTVCP 236  
 QY 265 MIDVIDHDFRYETQAG--DMRGAAPWEMYYKRIPIPP--ELQKADPSDPFSPVMAAG 320  
 DB 237 IIDVIDSDTPEY--MAGSDWTYGFNMKLNFRWYPPVQREMRRKGRDTPVPTWAGG 294  
 QY 321 LPAVDRKFMWELGMYDGLIWMGEQYETSFKTMWCGRMEDIPCISVGHIRYKRVYKXV 380  
 DB 295 LFSIDRYFOEITGTAGMDIMWGENLEISFRIMQCGTLEIYVCSVHGVFRKATPTTP 354  
 QY 381 PAGVVS--LARNLKRVAEVMWDEVAEYIYQRRPEYRHLISAGDVAVQKKLRSSLNCKSFKWF 438  
 DB 355 PGGTGQIINKNRRRLAEVWMDKFNFPYIISPEVTAKVDYDISRGLRKLQCKPFSWY 414  
 QY 415 LENIYPSQIPRYH-----FSLGEIRVETNQCLDNMARKEKNGKIFNC--HGNG--- 463  
 DB 439 MTKIAMD--LPKFPVPEPPAAAMGEIRNVGTGLCADTKHAGLSPLRLGCVGRGEAA 496  
 QY 497 WNNMVPFTPTWRDRIIRGDPQHTKKCFDAISITSPVTLVDCISMGMGNQAMKRXOK--TL 555  
 DB 464 --GNQVFSYANKKEIRDD-----LCLDVSKLNGVPTMLKCHRLKGNQIMBEPVYTL 515  
 QY 556 YHPVSGSCMD--CESDHRF--FMNTCNPSLSLQOMLFPHNTSYLLEK 600  
 DB 516 QHYNMNOCLDKATEBDSQVPSINDCN--GSRSQMWLR--NVTLPPLF 559







XX The present invention relates to nucleotide sequences which are useful in  
 CC the screening of compounds for the treatment of pain, or for the  
 CC diagnosis of pain. The nucleotide sequences are up-regulated in the  
 CC spinal cord in response to streptozocin-induced diabetes. The present  
 CC sequence was used to illustrate the invention.

XX Sequence 559 AA;

Query Match 34.0%; Score 1116; DB 7; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 2.3e-104;  
 Matches 230; Conservative 88; Mismatches 171; Indels 38; Gaps 16;

```

QY GNGEGRP--YPTMDARVDAVRENGFNIVSDKISLNSLPDIRHPCNSKRYLETLP 145
DB 57 GPGEMGRPVVLPKEDQKMKEMFKINQFNLMASBMAFNRLPVRLEGCTKYVPPSLP 116
QY 146 NTSLIIFPHNEGSSLLRTVSHVNLNSRPPBELVAEIVYVNDPSDRHLKKPLEDMY-ALFP 204
DB 117 TTSVIVFHNEMASTLLRTVSHVNLNSRPPBELVAEIVYVNDPSDRHLKKPLEDMY 176
QY 205 SVRLRTKKEGGLRTMLGASVATGVTITFLDSHCANVMPLPDLRIARNKTIYCP 264
DB 177 PVAHIRMGRSGILIRARLKGAAVSKGVITFLDAHCECTGWLPLARIKHDRATVYCP 236
QY 265 MIDVIDHDDFRYETQAG-DAMRGAFWEMYYKRIPIPP--ELQKADPSDPFESPVNAAG 320
DB 237 IIDVISDITFEY--MAGSDMTYGGFNWKLNFWRYPVQREMDRKGDRTLPVRTPTWAG 294
QY 321 LFAVDRKFWELGSDYDGLIENGEOYEISFKVMCGRMEDIPCSRVGHYRYKYVYK 380
DB 295 LFSIDRYFOEIGTYDAGMDIWGEMLEISFRIMQCGTLEIYCSHVGHVFRATPTYTF 354
QY 381 PAGVS--LAENLKRVAEVMDEVAEYIYORRPEYRHLASGVAVQKLRSLNCKSPKWF 438
DB 355 PGGTGQIINKNNRRLAEVWMDERKNFYIISPGVTKYVDYDISSRVLGRHLQCKPFSWY 414
QY 439 MTKIAMD--LPKFPYVPEPPAAAMGEIRNVGTGICADTKHAGLSPRLBGCVRGRBAA 496
DB 415 LENIYPSQIPIRHY-----FSLGEIRNVETNOCLDNMARKENKYGIFNC-HGMG--- 463
QY 497 WNNNQVFTFWREDIRPGDPQHTKKCFDAISHTSPVTLVYDCHSMKNQMLMKYRKDK-TL 555
DB 464 --GNQVFSYRANKKIRDD-----LCLDVSKLNGPVTMLKCHLKNQMLBEDPVKLT 515
QY 556 YHPVSGSCMD-CSESDEIRI-FMNTCNPSLSLQOQLFEHTNSTVLEK 600
DB 516 QHVNNSQCLDKATEBDSQVPSIRDCT-GSRSQMWLR--NVTLPPIF 559

```

RESULT 24

ADB79796  
 ID ADB79796 standard; protein; 559 AA.

XX ADB79796;

XX 04-DEC-2003 (first entry)

DE Rat polypeptide GalINAC transferase T1, SEQ ID 36.

KM Analgesic; pain; streptozocin-induced diabetes; rat.

OS Rattus norvegicus.

PN EPI279744-A2.

XX 29-JAN-2003.

PF 26-JUL-2002; 2002EP-00255249.

XX 27-JUL-2001; 2001GB-00018354.

PR 07-FEB-2002; 2002GB-00002910.

XX

PA (WARN ) WARNER LAMBERT CO.

XX Brooksbank RA, Dixon AK, Lee K, Pinnock RD;

XX WPI; 2003-395407/38.

DR N-PSDB; ADB79797.

XX Use of isolated gene sequences and encoded polypeptides that are  
 PT upregulated in the spinal cord in response to streptozocin-induced  
 PT diabetes for screening compounds for the treatment of pain, or for  
 PT diagnosing pain.

PS Claim 1; Page 91-93; 334pp; English.

XX The present invention relates to nucleotide sequences which are useful in  
 CC the screening of compounds for the treatment of pain, or for the  
 CC diagnosis of pain. The nucleotide sequences are up-regulated in the  
 CC spinal cord in response to streptozocin-induced diabetes. The present  
 CC sequence was used to illustrate the invention.

XX Sequence 559 AA;

Query Match 34.0%; Score 1116; DB 7; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 2.3e-104;  
 Matches 230; Conservative 88; Mismatches 171; Indels 38; Gaps 16;

```

QY GNGEGRP--YPTMDARVDAVRENGFNIVSDKISLNSLPDIRHPCNSKRYLETLP 145
DB 57 GPGEMGRPVVLPKEDQKMKEMFKINQFNLMASBMAFNRLPVRLEGCTKYVPPSLP 116
QY 146 NTSLIIFPHNEGSSLLRTVSHVNLNSRPPBELVAEIVYVNDPSDRHLKKPLEDMY-ALFP 204
DB 117 TTSVIVFHNEMASTLLRTVSHVNLNSRPPBELVAEIVYVNDPSDRHLKKPLEDMY 176
QY 205 SVRLRTKKEGGLRTMLGASVATGVTITFLDSHCANVMPLPDLRIARNKTIYCP 264
DB 177 PVAHIRMGRSGILIRARLKGAAVSKGVITFLDAHCECTGWLPLARIKHDRATVYCP 236
QY 265 MIDVIDHDDFRYETQAG-DAMRGAFWEMYYKRIPIPP--ELQKADPSDPFESPVNAAG 320
DB 237 IIDVISDITFEY--MAGSDMTYGGFNWKLNFWRYPVQREMDRKGDRTLPVRTPTWAG 294
QY 321 LFAVDRKFWELGSDYDGLIENGEOYEISFKVMCGRMEDIPCSRVGHYRYKYVYK 380
DB 295 LFSIDRYFOEIGTYDAGMDIWGEMLEISFRIMQCGTLEIYCSHVGHVFRATPTYTF 354
QY 381 PAGVS--LAENLKRVAEVMDEVAEYIYORRPEYRHLASGVAVQKLRSLNCKSPKWF 438
DB 355 PGGTGQIINKNNRRLAEVWMDERKNFYIISPGVTKYVDYDISSRVLGRHLQCKPFSWY 414
QY 439 MTKIAMD--LPKFPYVPEPPAAAMGEIRNVGTGICADTKHAGLSPRLBGCVRGRBAA 496
DB 415 LENIYPSQIPIRHY-----FSLGEIRNVETNOCLDNMARKENKYGIFNC-HGMG--- 463
QY 497 WNNNQVFTFWREDIRPGDPQHTKKCFDAISHTSPVTLVYDCHSMKNQMLMKYRKDK-TL 555
DB 464 --GNQVFSYRANKKIRDD-----LCLDVSKLNGPVTMLKCHLKNQMLBEDPVKLT 515
QY 556 YHPVSGSCMD-CSESDEIRI-FMNTCNPSLSLQOQLFEHTNSTVLEK 600
DB 516 QHVNNSQCLDKATEBDSQVPSIRDCT-GSRSQMWLR--NVTLPPIF 559

```

RESULT 25

ADB62526  
 ID ADB62526 standard; protein; 559 AA.

XX ADB62526;

XX 29-JAN-2004 (first entry)

DE Rat Protein Q10473, SEQ ID NO 8456.

XX



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 13:39:57, Search time 21 Seconds

(Without alignments)  
1904.274 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278

Sequence: 1 MRKKERKLQAVLVLAALV.....TQGMFPHNTSYLKEKRN 603

Scoring table: BIOSUM62

Gapop 10.0, Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*

2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*

3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*

4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*

5: /cgn2\_6/ptodata/1/1aa/PTCUS\_COMB.pep:\*

6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3278	100.0	603	4	US-09-795-926-43
2	3242.5	98.9	631	4	US-09-795-926-41
3	2771	84.5	506	4	US-09-795-926-31
4	2746.5	83.8	535	4	US-09-795-926-29
5	1877	57.3	366	4	US-09-795-926-39
6	1370	41.8	269	4	US-09-795-926-27
7	1311	40.0	321	4	US-09-795-926-35
8	1115	34.0	517	2	US-08-967-508-19
9	1115	34.0	517	3	US-08-967-506-19
10	1115	34.0	517	5	PCT-US94-02552-19
11	1115	34.0	559	2	US-08-967-508-9
12	1115	34.0	559	3	US-08-967-506-9
13	1115	34.0	559	5	PCT-US94-02552-9
14	1026	31.3	578	4	US-09-217-3068-2
15	1024	31.2	560	4	US-09-217-3068-10
16	990	30.2	240	4	US-09-795-926-33
17	988	30.1	209	4	US-09-795-926-37
18	971	29.6	552	4	US-10-140-002-196
19	941	28.7	302	4	US-09-376-856-2
20	926	28.2	639	3	US-09-376-856-2
21	920	28.1	638	3	US-09-347-488-2
22	916	27.9	633	2	US-08-648-298-2
23	804	24.5	224	4	US-09-795-926-23
24	744	22.7	603	4	US-09-973-457-2
25	675	20.6	289	4	US-09-270-767-45334
26	561	17.1	238	4	US-09-270-767-46411
27	503.5	15.4	366	4	US-09-270-767-43543

28	483	14.7	143	4	US-09-795-926-21	Sequence 21, Appl
29	481	14.7	112	4	US-09-795-926-25	Sequence 25, Appl
30	450.5	13.7	149	4	US-09-973-457-6	Sequence 6, Appl
31	428.5	13.1	161	4	US-09-270-767-58910	Sequence 58910, A
32	287	8.8	122	4	US-09-513-999C-7570	Sequence 7570, Ap
33	277.5	8.5	148	4	US-09-270-767-32872	Sequence 32872, A
34	241.5	7.4	313	4	US-09-270-767-46209	Sequence 46209, A
35	234.5	7.2	300	4	US-09-270-767-43809	Sequence 43809, A
36	228.5	7.0	290	4	US-09-270-767-61773	Sequence 61773, A
37	218.5	6.7	72	4	US-09-270-767-61989	Sequence 61989, A
38	168	5.1	73	4	US-09-270-767-33350	Sequence 33350, A
39	167	5.1	73	4	US-09-270-767-48567	Sequence 48567, A
40	167	5.1	101	4	US-09-270-767-33441	Sequence 33441, A
41	167	5.1	101	4	US-09-270-767-48568	Sequence 48568, A
42	133	4.1	424	4	US-09-134-000C-5947	Sequence 5947, Ap
43	129	3.9	88	4	US-09-270-767-33601	Sequence 33601, A
44	129	3.9	88	4	US-09-270-767-48818	Sequence 48818, A
45	128	3.9	135	4	US-09-973-457-5	Sequence 5, Appl
46	127	3.9	436	4	US-09-338-352-4416	Sequence 4416, Ap
47	125.5	3.8	965	4	US-09-437-277-3	Sequence 3, Appl
48	124	3.8	702	4	US-09-437-277-1	Sequence 1, Appl
49	123.5	3.8	534	4	US-09-134-000C-5087	Sequence 5087, Ap
50	123	3.8	721	4	US-09-107-532A-6889	Sequence 6889, Ap
51	121.5	3.7	1056	4	US-09-134-000C-5086	Sequence 5086, Ap
52	119.5	3.6	492	2	US-08-468-812-4	Sequence 4, Appl
53	119.5	3.6	492	3	US-08-468-812-7	Sequence 7, Appl
54	119.5	3.6	492	3	US-08-590-563-4	Sequence 4, Appl
55	119.5	3.6	492	3	US-08-590-563-7	Sequence 7, Appl
56	119.5	3.6	492	4	US-09-770-621-4	Sequence 4, Appl
57	119.5	3.6	492	4	US-09-770-621-7	Sequence 7, Appl
58	119.5	3.6	492	4	US-09-235-832-7	Sequence 7, Appl
59	119.5	3.6	492	4	US-09-235-832-4	Sequence 4, Appl
60	118	3.6	480	2	US-08-468-812-5	Sequence 5, Appl
61	118	3.6	480	3	US-08-590-563-5	Sequence 5, Appl
62	118	3.6	480	4	US-09-770-621-5	Sequence 5, Appl
63	118	3.6	480	4	US-09-235-832-5	Sequence 5, Appl
64	116.5	3.6	88	4	US-09-270-767-33890	Sequence 33890, A
65	116.5	3.6	88	4	US-09-270-767-49107	Sequence 49107, A
66	112.5	3.4	909	3	US-09-425-383-2	Sequence 2, Appl
67	111.5	3.4	491	2	US-08-468-812-8	Sequence 8, Appl
68	111.5	3.4	491	3	US-08-590-563-8	Sequence 8, Appl
69	111.5	3.4	491	3	US-09-770-621-8	Sequence 8, Appl
70	111.5	3.4	491	4	US-09-235-832-8	Sequence 8, Appl
71	107.5	3.3	507	4	US-09-130-337A-25	Sequence 25, Appl
72	107	3.3	270	4	US-09-495-406-25	Sequence 25, Appl
73	107	3.3	270	4	US-09-816-028A-39	Sequence 39, Appl
74	107	3.3	270	4	US-10-303-162-39	Sequence 39, Appl
75	105.5	3.2	187	4	US-09-973-457-4	Sequence 4, Appl
76	105.5	3.2	336	4	US-09-543-681A-6535	Sequence 6535, Ap
77	105	3.2	127	1	US-08-392-828C-39	Sequence 39, Appl
78	105	3.2	127	3	US-09-330-945-39	Sequence 39, Appl
79	105	3.2	127	3	US-08-597-236-10	Sequence 10, Appl
80	105	3.2	324	1	US-08-746-682A-10	Sequence 10, Appl
81	102.5	3.1	540	1	US-08-378-761A-77	Sequence 77, Appl
82	102.5	3.1	540	1	US-08-485-286-77	Sequence 77, Appl
83	102	3.1	317	4	US-09-583-110-4849	Sequence 4849, Ap
84	101.5	3.1	395	4	US-08-635-552A-4	Sequence 4, Appl
85	101.5	3.1	419	2	US-08-270-581-2	Sequence 2, Appl
86	101.5	3.1	419	2	US-09-146-893-2	Sequence 2, Appl
87	101.5	3.1	419	4	US-08-675-499A-5	Sequence 5, Appl
88	101.5	3.1	419	4	US-08-812-008-5	Sequence 5, Appl
89	101	3.1	328	4	US-09-583-110-3554	Sequence 3554, Ap
90	100	3.1	331	4	US-09-634-238-47	Sequence 247, Ap
91	99.5	3.0	303	4	US-09-495-406-17	Sequence 17, Appl
92	99.5	3.0	303	4	US-09-816-028A-29	Sequence 29, Appl
93	99.5	3.0	303	4	US-10-303-162-29	Sequence 29, Appl
94	99	3.0	341	4	US-09-328-352-1178	Sequence 7178, Ap
95	99	3.0	301	4	US-09-495-406-15	Sequence 15, Appl
96	98	3.0	301	4	US-09-816-028A-27	Sequence 27, Appl
97	98	3.0	301	4	US-10-303-162-27	Sequence 27, Appl
98	96.5	2.9	623	1	US-08-653-740-7	Sequence 7, Appl
99	96.5	2.9	623	2	US-09-073-594-7	Sequence 7, Appl
100	96.5	2.9	623	3	US-09-275-925-7	Sequence 7, Appl

## ALIGNMENTS

## RESULT 1

US-09-795-926-43  
Sequence 43, Application US/09795926  
Patent No. 6555669  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abuhin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kiele, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME  
CURRENT FILING DATE: 2001-02-28  
CURRENT APPLICATION NUMBER: US/09/795,926  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 43  
LENGTH: 603  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-43

Query Match 100.0%; Score 3278; DB 4; Length 603;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MRRREKLLQAVLVLAVALVLLPVGGLMALYRERQPDGTGSGAATAVPAAGQSHSROK 60  
DB 1 MRRREKRLQAVLVLAVALVLLPVGGLMALYRERQPDGTGSGAATAVPAAGQSHSROK 60  
QY 61 KTFELGDSQKLDKMDHDEKAIIRDAQVRVNGEGRPYPMTDARVDQAYRENGFNIVYSDK 120  
DB 61 KTFELGDSQKLDKMDHDEKAIIRDAQVRVNGEGRPYPMTDARVDQAYRENGFNIVYSDK 120  
QY 121 ISLRSLSPLDIHPNCNKRILETLPTNTSIIIPFNHSGSSILRTVHSLVNSPPELVAEI 180  
DB 121 ISLRSLSPLDIHPNCNKRILETLPTNTSIIIPFNHSGSSILRTVHSLVNSPPELVAEI 180  
QY 181 VLVDPSRREHLKKPLLEDYMLPFSVRILRTKKREGILRTMLGASVATGVTITFLDSHC 240  
DB 181 VLVDPSRREHLKKPLLEDYMLPFSVRILRTKKREGILRTMLGASVATGVTITFLDSHC 240  
QY 241 EANNVMLPPLLDRLIARNKRTIVCPMIDVIDHDDFRYETQAGDAMRGAPDMYKRIPIIP 300  
DB 241 EANNVMLPPLLDRLIARNKRTIVCPMIDVIDHDDFRYETQAGDAMRGAPDMYKRIPIIP 300  
QY 301 PELQADSDPFPESPVMAAGLPAVDKRMFWELGGYDPLGLIWGSGQYEISFKVMWCGGRM 360  
DB 301 PELQADSDPFPESPVMAAGLPAVDKRMFWELGGYDPLGLIWGSGQYEISFKVMWCGGRM 360  
QY 361 EDIPCSRGHTYRKVPYKVPAGVSLANTKRVAAVMMDEYAEIYYQRRPSTYRLHSADV 420  
DB 361 EDIPCSRGHTYRKVPYKVPAGVSLANTKRVAAVMMDEYAEIYYQRRPSTYRLHSADV 420

## RESULT 2

US-09-795-926-41  
Sequence 41, Application US/09795926  
Patent No. 6555669  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abuhin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kiele, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME  
CURRENT FILING DATE: 2001-02-28  
CURRENT APPLICATION NUMBER: US/09/795,926  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 41  
LENGTH: 631  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-41

Query Match 98.9%; Score 3242.5; DB 4; Length 631;  
Best Local Similarity 95.4%; Pred. No. 0;  
Matches 601; Conservative 0; Mismatches 29; Indels 1;  
QY 1 MRRREKLLQAVLVLAVALVLLPVGGLMALYRERQPDGTGSGAATAVPAAGQSHSROK 60  
DB 1 MRRREKRLQAVLVLAVALVLLPVGGLMALYRERQPDGTGSGAATAVPAAGQSHSROK 60  
QY 61 KTFELGDSQKLDKMDHDEKAIIRDAQVRVNGEGRPYPMTDARVDQAYRENGFNIVYSDK 120  
DB 61 KTFELGDSQKLDKMDHDEKAIIRDAQVRVNGEGRPYPMTDARVDQAYRENGFNIVYSDK 120  
QY 121 ISLRSLSPLDIHPNCNKRILETLPTNTSIIIPFNHSGSSILRTVHSLVNSPPELVAEI 180  
DB 121 ISLRSLSPLDIHPNCNKRILETLPTNTSIIIPFNHSGSSILRTVHSLVNSPPELVAEI 180  
QY 181 VLVDPSRREHLKKPLLEDYMLPFSVRILRTKKREGILRTMLGASVATGVTITFLDSHC 240  
DB 181 VLVDPSRREHLKKPLLEDYMLPFSVRILRTKKREGILRTMLGASVATGVTITFLDSHC 240

QY 241 EAVNVMPLPDLRIRARKTIVCPMIDVIDHDDFRYENQAGDAMRGAFDWEMYKRIPIR 300  
DB 241 EAVNVMPLPDLRIRARKTIVCPMIDVIDHDDFRYENQAGDAMRGAFDWEMYKRIPIR 300  
QY 301 PELQKADSPDFESPVAAGGLFAVDRKMFELAGYDQLEIWGEQOYEISFK----- 352  
DB 301 PELQKADSPDFESPVAAGGLFAVDRKMFELAGYDQLEIWGEQOYEISFKGLHMLPRL 360  
QY 353 -----VMMCGRMEDIPCSRGHYIRKTYPKVPAGVSLARNLK 391  
DB 361 VSNMGPQAVFLPAPNMLAQVMMCGRMEDIPCSRGHYIRKTYPKVPAGVSLARNLK 420  
QY 392 RVLEVMWDEYAEYIYORRPEYRHLISAGDVAVQKKLRSLNCKSPKFMPTKIAMDLPKFYR 451  
DB 421 RVLEVMWDEYAEYIYORRPEYRHLISAGDVAVQKKLRSLNCKSPKFMPTKIAMDLPKFYR 480  
QY 452 PVBPAAAMGSEIRNVGTGLCADTKHGALGSPRLLEGCVRGGEAANNMNVFTTWREDI 511  
DB 481 PVBPAAAMGSEIRNVGTGLCADTKHGALGSPRLLEGCVRGGEAANNMNVFTTWREDI 540  
QY 512 RPQDPQHTKKPCFPAISHTSPVTLVDCHSMKGNQLMKTKRDKTLVHPVSGSCMDSSSDH 571  
DB 541 RPQDPQHTKKPCFPAISHTSPVTLVDCHSMKGNQLMKTKRDKTLVHPVSGSCMDSSSDH 600  
QY 572 RIFMNTCPSSLTQOOWLFEHTNSTVLEKFN 601  
DB 601 RIFMNTCPSSLTQOOWLFEHTNSTVLEKFN 630

## RESULT 3

US-09-795-926-31  
Sequence 31, Application US/09795926  
Patent No. 6555669  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abuin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kieke, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: LEX-0144-USA  
CURRENT APPLICATION NUMBER: US/09/795, 926  
CURRENT FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186,558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 31  
LENGTH: 506  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-31

Query Match 84.5%; Score 2771; DB 4; Length 506;

Best Local Similarity 100.0%; Pred. No. 2.7e-289; Indels 0; Gaps 0;

QY 98 MTDARVDQAVRENGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 157  
DB 1 MTDARVDQAVRENGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 60

QY 158 WSSLRTVHSLVLRSPPELVAEIVLVDDFSDBREHLKKPLLEDYMLPPSVRIILRTKKEGL 217  
DB 61 WSSLRTVHSLVLRSPPELVAEIVLVDDFSDBREHLKKPLLEDYMLPPSVRIILRTKKEGL 120  
QY 218 IRTMGLASVATGVTITFLDSHCANVMVLPDLRIARRKTIVCPMIDVIDHDDFRYE 277  
DB 121 IRTMGLASVATGVTITFLDSHCANVMVLPDLRIARRKTIVCPMIDVIDHDDFRYE 180  
QY 278 TQAGDAMRGAFDWEMYKRIPIPELOKADSPDFESPVAAGGLFAVDRKMFELAGYD 337  
DB 181 TQAGDAMRGAFDWEMYKRIPIPELOKADSPDFESPVAAGGLFAVDRKMFELAGYD 240  
QY 338 GLEIWGEQOYEISFKVMMCGRMEDIPCSRGHYIRKTYPKVPAGVSLARNLKRVAEVW 397  
DB 241 GLEIWGEQOYEISFKVMMCGRMEDIPCSRGHYIRKTYPKVPAGVSLARNLKRVAEVW 300  
QY 398 MDEYAEYIYORRPEYRHLISAGDVAVQKKLRSLNCKSPKFMPTKIAMDLPKFYR 457  
DB 301 MDEYAEYIYORRPEYRHLISAGDVAVQKKLRSLNCKSPKFMPTKIAMDLPKFYR 360  
QY 458 AAMGEIRNVGTGLCADTKHGALGSPRLLEGCVRGGEAANNMNVFTTWREDIRPQDP 517  
DB 361 AAMGEIRNVGTGLCADTKHGALGSPRLLEGCVRGGEAANNMNVFTTWREDIRPQDP 420  
QY 518 HTKKPCFPAISHTSPVTLVDCHSMKGNQLMKTKRDKTLVHPVSGSCMDSSSDHIFMNT 577  
DB 421 HTKKPCFPAISHTSPVTLVDCHSMKGNQLMKTKRDKTLVHPVSGSCMDSSSDHIFMNT 480  
QY 578 CNPSSLTQOOWLFEHTNSTVLEKFN 603  
DB 481 CNPSSLTQOOWLFEHTNSTVLEKFN 506

## RESULT 4

US-09-795-926-29  
Sequence 29, Application US/09795926  
Patent No. 6555669  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abuin, Alejandro  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kieke, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: LEX-0144-USA  
CURRENT APPLICATION NUMBER: US/09/795, 926  
CURRENT FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186,558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 29  
LENGTH: 535  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-29

Query Match 83.8%; Score 2746.5; DB 4; Length 535;

Best Local Similarity 94.6%; Pred. No. 1.3e-286; Indels 29; Gaps 1;

```
Qy 98 MTDARVDOAYRENGFNIIYSDKISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNEG 157
Db 1 MTDARVDOAYRENGFNIIYSDKISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNEG 60
Qy 158 WSSILRTVHSVLRNSPELVAEIVLVDPSDRHLKKPLEDYMALPFSVRILRTKKEGL 217
Db 61 WSSILRTVHSVLRNSPELVAEIVLVDPSDRHLKKPLEDYMALPFSVRILRTKKEGL 120
Qy 218 IRTMGLASVATGVITFLDSHCANVMVLPILLDIRARNKTIVCPIIDVIDHDDPRYE 277
Db 121 IRTMGLASVATGVITFLDSHCANVMVLPILLDIRARNKTIVCPIIDVIDHDDPRYE 180
Qy 278 TOAGDARGAFFDMWYKRIPIPELOKADSPDFESFVMAAGLFAVDRKMFELGGYDP 337
Db 181 TOAGDARGAFFDMWYKRIPIPELOKADSPDFESFVMAAGLFAVDRKMFELGGYDP 240
Qy 338 GLEIWGEBOYEISFK-----VMMCGMEDIPCSRV 368
Db 241 GLEIWGEBOYEISFKGLHMLPRLVNSWVQAFLPRAFMALQVMMCGMEDIPCSRV 300
Qy 369 GHIRKXVVPYKPVAGVSLARBLKXVAEYVMDVAYEYIQRRPEYRHLISAGDVAOKLRS 428
Db 301 GHIRKXVVPYKPVAGVSLARBLKXVAEYVMDVAYEYIQRRPEYRHLISAGDVAOKLRS 360
Qy 429 SLNCKSFKFWMTKIAMDLPKFYPPVBPAAWGEIRNVGTGLCADTKHGLSGPLRLEG 488
Db 361 SLNCKSFKFWMTKIAMDLPKFYPPVBPAAWGEIRNVGTGLCADTKHGLSGPLRLEG 420
Qy 489 VRGGEAAMNMVFTFTWRDIPRGPOTTKKCFDPAISHTSPVTLTDCHSMKGNOLWK 548
Db 421 VRGGEAAMNMVFTFTWRDIPRGPOTTKKCFDPAISHTSPVTLTDCHSMKGNOLWK 480
Qy 549 YRKXKTYHPVSGSCMDCSDHRIFMNTCNPSLTOOMLEHNTSTLEKFNEN 603
Db 481 YRKXKTYHPVSGSCMDCSDHRIFMNTCNPSLTOOMLEHNTSTLEKFNEN 535

RESULT 5
US-09-795-926-39
; Sequence 39, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Milganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 366
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-39
```

```
Query Match 57.3%; Score 1877; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 3,5e-193;
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRRKREKLLDAVALVALVLAALVLLPVGGMALYRERQPTGGSGAIVAPAGGSHSRQK 60
Db 1 MRRKREKLLDAVALVALVLAALVLLPVGGMALYRERQPTGGSGAIVAPAGGSHSRQK 60
Qy 61 KTFPLGSGOKLQKMDHDEA1RRDAQRVNGEGQGRPYMTAERDOAYRENGFNIIYSDK 120
Db 61 KTFPLGSGOKLQKMDHDEA1RRDAQRVNGEGQGRPYMTAERDOAYRENGFNIIYSDK 120
Qy 121 ISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNEGSSILRTVHSVLRNSPELVAEI 180
Db 121 ISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNEGSSILRTVHSVLRNSPELVAEI 180
Qy 181 VLVDPSDRHLKKPLEDYMALPFSVRILRTKKEGLIRTRMLGASVATGDTVITFLSHC 240
Db 181 VLVDPSDRHLKKPLEDYMALPFSVRILRTKKEGLIRTRMLGASVATGDTVITFLSHC 240
Qy 241 EANNVNLPLLDRIARNRKTIVCPIIDVIDHDDPRYETOAGDARGAFFDMWYKRIPI 300
Db 241 EANNVNLPLLDRIARNRKTIVCPIIDVIDHDDPRYETOAGDARGAFFDMWYKRIPI 300
Qy 301 PELQKADSPDFESFVMAAGLFAVDRKMFELGGYDGLIWGGEBOYEISFKV 353
Db 301 PELQKADSPDFESFVMAAGLFAVDRKMFELGGYDGLIWGGEBOYEISFKV 353

RESULT 6
US-09-795-926-27
; Sequence 27, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Milganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 269
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-27

Query Match 41.8%; Score 1370; DB 4; Length 269;
Best Local Similarity 100.0%; Pred. No. 9.1e-139;
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 98 MTDARVDOAYRENGFNIIYSDKISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNEG 157
Db 1 MTDARVDOAYRENGFNIIYSDKISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFNEG 60
```

Qy	158	WSLLKTVSHVLSNRSPPELVAAIYLVDPDSDEHKKELEDMYMLPESVRLITPKKREG	217
Db	61	WSLLKTVSHVLSNRSPPELVAAIYLVDPDSDEHKKELEDMYMLPESVRLITPKKREG	120
Qy	218	IRTRMLGASVATGDTVTPLDISHCEANVMMLPPLDLRIARNKRTIVCPMIDVIDHDDPRYE	277
Db	121	IRTRMLGASVATGDTVTPLDISHCEANVMMLPPLDLRIARNKRTIVCPMIDVIDHDDPRYE	160
Qy	278	TQNGDANRGAFDMEMTYKKRIPIPIPELQKADPSDPFESVVMAGGLFANDRKKFMELGSDP	337
Db	181	TQNGDANRGAFDMEMTYKKRIPIPIPELQKADPSDPFESVVMAGGLFANDRKKFMELGSDP	240
Qy	338	GLEIMGGEQYEISFKV	353
Db	241	GLEIMGGEQYEISFKV	256

## RESULT 7

```

US-09-795-926-35
? Sequence 35, Application US/09795926
? Patent No. 6555669
? GENERAL INFORMATION:
? APPLICANT: Donoho, Gregory
? APPLICANT: Hilbun, Erin
? APPLICANT: Turner, C. Alexander Jr.
? APPLICANT: Friedrich, Glenn
? APPLICANT: Abuin, Alejandro
? APPLICANT: Zambrowicz, Brian
? APPLICANT: Sands, Arthur T.
? APPLICANT: Walke, D. Wade
? APPLICANT: Milgnowski, Nathaniel L.
? APPLICANT: Hu, Yi
? APPLICANT: Kieke, James Alvin
? APPLICANT: Poter, David George
? TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
? FILE REFERENCE: LEX-0144-USA
? CURRENT APPLICATION NUMBER: US/09/795,926
? PRIOR FILING DATE: 2001-02-28
? PRIOR APPLICATION NUMBER: US 60/185,920
? PRIOR FILING DATE: 2000-02-29
? PRIOR APPLICATION NUMBER: US 60/186,558
? PRIOR FILING DATE: 2000-03-02
? PRIOR APPLICATION NUMBER: US 60/191,849
? PRIOR FILING DATE: 2000-03-24
? NUMBER OF SEQ ID NOS: 47
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 35
? LENGTH: 321
? TYPE: PRT
? ORGANISM: homo sapiens
US-09-795-926-35

```

	Query Match	40.0%;	Score 1311;	DB 4;	Length 321;
	Best Local Similarity	99.2%;	Pred. No. 2,8e-132;		
	Matches	251;	Conservative	1;	Mismatches 1; Indels 0; Gaps 0
QY	1	MRRREKLLQAAVLAALVLAALVLLPNVGLMALRYEROPDGTGGSGAAVPAAGQSGHSRQK	60		
Db	1	MRRREKRLQAAVLAALVLAALVLLPNVGLMALRYEROPDGTGGSGAAVPAAGQSGHSRQK	60		
QY	61	KTFPLGDOQKAKDMHDKAIRRDQAVNGEGQGRYPMTDARVVOAARENGFNLYYSDK	120		
Db	61	KTFPLGDOQKAKDMHDKAIRRDQAVNGEGQGRYPMTDARVVOAARENGFNLYYSDK	120		
QY	121	ISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFHNEGSSLLRTVHSLYLNRSPPELVAEI	180		
Db	121	ISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFHNEGSSLLRTVHSLYLNRSPPELVAEI	180		
QY	181	VLVDDFSRHLKKPLIEDYMALPESVRLLRTPKRGILRTMYLGSVATGDTITLDSHC	240		
Db	181	VLVDDFSRHLKKPLIEDYMALPESVRLLRTPKRGILRTMYLGSVATGDTITLDSHC	240		

```
QY      241 EAAVNWLPPLDR 253
          |||||
Db      : 241 EAAVNWLPPLGK 253
```

## RESULT

US-08-967-508-19  
Sequence 19, Application US/08967508  
Patent No. 5910570  
GENERAL INFORMATION:  
APPLICANT: The Upjohn Company  
APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.  
TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAC:  
NUMBER OF INVENTION: Polypeptide, N-Acetylgalactosaminyltransferase  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
ADDRESSER: Property Legal Services  
STREET: 301 Henrietta Street  
CITY: Kalamazoo  
STATE: Michigan  
COUNTRY: USA  
ZIP: 49001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,508  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,830  
FILING DATE: 13 No. 5910570ember 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Darnley Jr., James D.  
REGISTRATION NUMBER: 33,673  
REFERENCE/DOCKET NUMBER: 4755.P CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 616-833-2210  
TELEFAX: 616-833-8897  
TELEX: 224401  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 517 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-967-508-19

[illegible]



Db 253 LFSIDRYFGEIGYDAGMDIMGGENLEISFRIMQCGTLEIVTCSHVGHFRKATPYTP 312  
Qy 381 PAGVS--LARNLKVAEVMDEYAEYIYQRRPEYRHLSAGDVAVQKLRSLNCKSPKWF 438  
Db 313 PGGTQILINKNNRRLAEVWMDPEFNFFYIISPVTKYDYGDISRLGKHLKQCRPSWY 372  
Qy 439 MTKIAMD--LPKFPVPPEPAAWGEIRNVGTGLCADTKHGALSPRLBGCVRGGEAA 496  
Db 373 LENIYPSQIPRHY-----FSLGEIRNVETNOCLDMARKENKVGIFNC-HGMG--- 421  
Qy 497 WNNQVFTFTREDIRPDPQHTKKFCFDAISHTSPVTLTYDCHSMKGNQMLKRYKDK-TL 555  
Db 422 --GNQVFSYTNKKEIRTD-----LCIDVSKLNGPVTMLKCHHLKGNQMLBEPVLTLL 473  
Qy 556 YHVPVSGCMD-CSESDEHRI-FMNTCNPSLSLTOQMLFHTNSTVLEKF 600  
Db 474 QHVNNOCLDKATBEDSQVPSIRDCS-GSRSQOMLLR--NVTLEPIF 517

RESULT 9  
US-08-967-506-19

Sequence 19, Application US/08967506  
Patent No. 6096512  
GENERAL INFORMATION:  
APPLICANT: The Upjohn Company  
APPLICANT: FOR U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.  
TITLE OF INVENTION: A cloned DNA Encoding a UDP-GALNAC:  
TITLE OF INVENTION: Polypeptide, N-Acetylglactosaminyltransferase  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
ADDRESSEE: Property Legal Services  
STREET: 301 Henrietta Street  
CITY: Kalamazoo  
STATE: Michigan  
COUNTRY: USA  
ZIP: 49001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,506  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,830  
FILING DATE: 13 NO. 6096512ember 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Darnley Jr., James D.  
REGISTRATION NUMBER: 33,673  
REFERENCE/DOCKET NUMBER: 4755.P CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 616-833-2210  
TELEFAX: 616-833-8897  
TELEX: 224401  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 517 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-967-506-19

Query Match 34.0%; Score 1115; DB 3; Length 517;  
Best Local Similarity 43.6%; Pred. No. 8,3e-111;  
Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;  
Qy 88 GNGEGRP--YPMDDAENVDAVRENGFNIVYSDKISLNRSLPIDIRHNCNSKRYLETLP 145

Db 15 GPGEMGKPVVIPKEDQKMKEMKINOFLNLMASEMLALNRLSPDVRLEGCKTKYYPDNL 74  
Qy 146 NTSIIIPFHNESGSLRTVHSVNLNRPPELVAVETLVDDPSREHKKPLEEDYM-ALFP 204  
Db 75 TTSVIVIFHNHASTLTARTHSVNRSPRMLBEIVLVDDASBDFLAKRLESVKKLV 134  
Qy 205 SVRIIRTKKKBGLIRITMLGASVATGVTITFLDSHCANVNLPLLDRIARNKTVCP 264  
Db 135 PVHVRMEQSRSGLIRALKGAAVSKQGVITFLDAHCECTGVMEPLARIKHDRKTVCP 194  
Qy 265 MIDVIDHDDFRYERQAG-DAMRGAEDMEYKKRIPIPE---ELQKADSPDFESSPVNAG 320  
Db 195 IIVDISDTEY--MAQSDMTYGGFNKKLFRMYPVQREMDRKGDRITLVRTPTNAGG 252  
Qy 321 LFAVDRKFMELAGYDPEGLIEWGEQYISFKVMCCGMRMEDIPCSRVGHYIRKRYPVYK 380  
Db 253 LFSIDRYFGEIGYDAGMDIMGGENLEISFRIMQCGTLEIVTCSHVGHFRKATPYTP 312  
Qy 381 PAGVS--LARNLKVAEVMDEYAEYIYQRRPEYRHLSAGDVAVQKLRSLNCKSPKWF 438  
Db 313 PGGTQILINKNNRRLAEVWMDPEFNFFYIISPVTKYDYGDISRLGKHLKQCRPSWY 372  
Qy 439 MTKIAMD--LPKFPVPPEPAAWGEIRNVGTGLCADTKHGALSPRLBGCVRGGEAA 496  
Db 373 LENIYPSQIPRHY-----FSLGEIRNVETNOCLDMARKENKVGIFNC-HGMG--- 421  
Qy 497 WNNQVFTFTREDIRPDPQHTKKFCFDAISHTSPVTLTYDCHSMKGNQMLKRYKDK-TL 555  
Db 422 --GNQVFSYTNKKEIRTD-----LCIDVSKLNGPVTMLKCHHLKGNQMLBEPVLTLL 473  
Qy 556 YHVPVSGCMD-CSESDEHRI-FMNTCNPSLSLTOQMLFHTNSTVLEKF 600  
Db 474 QHVNNOCLDKATBEDSQVPSIRDCS-GSRSQOMLLR--NVTLEPIF 517

RESULT 10  
PCT-US94-02552-19

Sequence 19, Application PC/TUS9402552  
GENERAL INFORMATION:  
APPLICANT: Elhammer, Ake P.  
APPLICANT: Homa, Fred L.  
TITLE OF INVENTION: A cloned DNA Encoding a UDP-GALNAC:  
TITLE OF INVENTION: Polypeptide, N-Acetylglactosaminyltransferase  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The Upjohn Company, Corp. Intellectual  
ADDRESSEE: Property Law  
STREET: 301 Henrietta Street  
CITY: Kalamazoo  
STATE: Michigan  
COUNTRY: USA  
ZIP: 49001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/02552  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Darnley Jr., James D.  
REGISTRATION NUMBER: 33,673  
REFERENCE/DOCKET NUMBER: 4755.P CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 616-385-5210  
TELEFAX: 616-385-6897  
TELEX: 224401  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 517 amino acids  
TYPE: amino acid



STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-02552-19

Query Match 34.0%; Score 115; DB 5; Length 517;  
Best Local Similarity 43.6%; Pred. No. 8.3e-111;  
Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGEQRP--YPTDAERVDQAYRENGFNIVYSDKISLNSLPDIRHPCNSKRYLETLP 145  
DB 15 GPGEMGRPVVLPKEDQKMKEMFKINQFNLMASEMIALNLSLPDVRLEGCKTKVYPNLP 74  
QY 146 NTSIIIPFHNCGSSLLRTVSHVSNRSPPELVAVIVDDPSDEHKKPEEDYM-ALFP 204  
DB 75 TTSVIVFHNEMASTLRTVSHVSNRSPPELVAVIVDDPSDEHKKPEEDYM-ALFP 134  
QY 205 SVRIATKKEGLIRTRMLGASVATGVTPLDSHCENAVNMLPPLDIRIARNKTIYCP 264  
DB 135 PVHVIEMGRSGILIRALKGAASVKGQVITFLDAHCECTGWLPLARIKHDRKTYVCP 194  
QY 265 MIDVIDHDDFRYETQAG-DAMGAFDWMYKRIPIPP--ELQKADPSDFESSPVNAG 320  
DB 195 IIDVISDITFEY--MAGSDMTYGGFNKMLNFRMYPVQREMDRKGRTLPVPTPTMAGG 252  
QY 321 LFAVDRKFMELGSDYDGLIHWGEOYEISFKVMCGRMEDIPCSRVGHYRKTYVYK 380  
DB 253 LFSIDRDFQELIGYDAGMDIWGENLEISFRIMOCGGTLEIYVCSHVGHVFRATPTYP 312  
QY 381 PAGVS--LARNLKVAVEMDEYAEIYQRRPEYRHLISAGDVAVOKKLRSLSNCKSPKWF 438  
DB 313 PGCGQIINNNRRLAEVWMDDEKMFYIISPGYTKVDYDGISSRLGIRHKLQCRPFSWY 372  
QY 439 MTKIAMD--LPKYPPEVPPAAWGEIRNVGTGLCADTKHAGLSPLRLGCVRGREAA 496  
DB 373 LENIYPSQIPRYH-----FSLGEIRNVETNOCLDNMARKENKVGIFNC-HGMG--- 421  
QY 497 WNNQVFTFMRDEIRPDQHTKKFCFDAISHTSPVTLVYDCHSKMGNOQMLKRYKDK-TL 555  
DB 422 --GNQVSYTANKKIRDD-----LCLDYSKANGPVTMLKCHHKGNOQMLBEPVLT 473  
QY 556 YHVPVSGCMD-CSESDHRI-FMNTCNPSLSLTOQWLFETHNSTYLEKF 600  
DB 474 QHVNNSQCLDKATDEDSQVPSIRDGS-GSRSQQWILR--NVTLPETIF 517

## RESULT 11

US-08-967-508-9  
Sequence 9, Application US/08967508  
Patent No. 5910570  
GENERAL INFORMATION:  
APPLICANT: The Upjohn Company  
APPLICANT: FOR U.S. PURPOSES ONLY: Elihammer, Ake P. and Homa, Fred L.  
TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GALNAC:  
TITLE OF INVENTION: Polypeptide, N-Acetylglactosaminyltransferase  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
STREET: 301 Henrietta Street  
CITY: Kalamazoo  
STATE: Michigan  
COUNTRY: USA  
ZIP: 49001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,508  
FILING DATE:  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pharmacia & Upjohn Company, Intellectual  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/602,830  
FILING DATE: 13 No. 5910570ember 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Darnley Jr., James D.  
REGISTRATION NUMBER: 33,673  
REFERENCE/DOCKET NUMBER: 4755.P CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 616-833-2210  
TELEFAX: 616-833-8897  
TELEX: 224401  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 559 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-967-508-9

Query Match 34.0%; Score 115; DB 2; Length 559;  
Best Local Similarity 43.6%; Pred. No. 9.5e-111;  
Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGEQRP--YPTDAERVDQAYRENGFNIVYSDKISLNSLPDIRHPCNSKRYLETLP 145  
DB 57 GPGEMGRPVVLPKEDQKMKEMFKINQFNLMASEMIALNLSLPDVRLEGCKTKVYPNLP 116  
QY 146 NTSIIIPFHNCGSSLLRTVSHVSNRSPPELVAVIVDDPSDEHKKPEEDYM-ALFP 204  
DB 117 TTSVIVFHNEMASTLRTVSHVSNRSPPELVAVIVDDPSDEHKKPEEDYM-ALFP 176  
QY 205 SVRIATKKEGLIRTRMLGASVATGVTPLDSHCENAVNMLPPLDIRIARNKTIYCP 264  
DB 177 PVHVIEMGRSGILIRALKGAASVKGQVITFLDAHCECTGWLPLARIKHDRKTYVCP 236  
QY 265 MIDVIDHDDFRYETQAG-DAMGAFDWMYKRIPIPP--ELQKADPSDFESSPVNAG 320  
DB 237 IIDVISDITFEY--MAGSDMTYGGFNKMLNFRMYPVQREMDRKGRTLPVPTPTMAGG 294  
QY 497 WNNQVFTFMRDEIRPDQHTKKFCFDAISHTSPVTLVYDCHSKMGNOQMLKRYKDK-TL 555  
DB 295 LFSIDRDFQELIGYDAGMDIWGENLEISFRIMOCGGTLEIYVCSHVGHVFRATPTYP 354  
QY 381 PAGVS--LARNLKVAVEMDEYAEIYQRRPEYRHLISAGDVAVOKKLRSLSNCKSPKWF 438  
DB 355 PGCGQIINNNRRLAEVWMDDEKMFYIISPGYTKVDYDGISSRLGIRHKLQCRPFSWY 414  
QY 439 MTKIAMD--LPKYPPEVPPAAWGEIRNVGTGLCADTKHAGLSPLRLGCVRGREAA 496  
DB 415 LENIYPSQIPRYH-----FSLGEIRNVETNOCLDNMARKENKVGIFNC-HGMG--- 463  
QY 497 WNNQVFTFMRDEIRPDQHTKKFCFDAISHTSPVTLVYDCHSKMGNOQMLKRYKDK-TL 555  
DB 464 --GNQVSYTANKKIRDD-----LCLDYSKANGPVTMLKCHHKGNOQMLBEPVLT 515  
QY 556 YHVPVSGCMD-CSESDHRI-FMNTCNPSLSLTOQWLFETHNSTYLEKF 600  
DB 516 QHVNNSQCLDKATDEDSQVPSIRDGS-GSRSQQWILR--NVTLPETIF 559

## RESULT 12

US-08-967-506-9  
Sequence 9, Application US/08967506  
Patent No. 6096512  
GENERAL INFORMATION:  
APPLICANT: The Upjohn Company  
APPLICANT: FOR U.S. PURPOSES ONLY: Elihammer, Ake P. and Homa, Fred L.  
TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GALNAC:  
TITLE OF INVENTION: Polypeptide, N-Acetylglactosaminyltransferase  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pharmacia & Upjohn Company, Intellectual



QY 439 MTKIAMD--LPKFPYVBPAAAMGEIRNVGTG:CADTKHAGLSPLRLBGCVRGREA 496  
DB 415 LENIYPSQJPRHY-----FSLGEIRNVETNOCLDMARKENKGEIFENC-HGWG-- 463  
QY 497 MNNOVFTFWREDIRPDPQHTKKFCFDAISHTSPVTLVYDCHSMKNQMLKRYKDK-TL 555  
DB 464 --GNQVFSYTNKKIRRDD-----LCLDVSKLNGPVTMLKCHLKNQMLMEYVPVLT 515  
QY 556 YHPVSGSCMD-CSESDBRI-FMNTCNBSLTQOMLFHTNSTVLKFP 600  
DB 516 QHVNNOCLDKATDEDSQVPSIRDCS-GSRQOMLIR--NVTLPEIF 559

RESULT 14  
US-09-217-306B-2  
Sequence 2, Application US/09217306B  
Patent No. 6465220  
GENERAL INFORMATION:  
APPLICANT: Haasan, Helle  
APPLICANT: Clausen, Henrik  
TITLE OF INVENTION: Glycosylation Using Galnac-T4 Transferase  
FILE REFERENCE: 8850\*1  
CURRENT APPLICATION NUMBER: US/09/217,306B  
CURRENT FILING DATE: 1998-12-21  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 2  
LENGTH: 578  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-217-306B-2

Query Match 31.3%; Score 1026; DB 4; Length 578;  
Best Local Similarity 39.4%; Pred. No. 3.9e-101;  
Matches 221; Conservative 98; Mismatches 204; Indels 38; Gaps 16;

QY 50 AAGGSSSRQKTTFLDGGQTKMDHKEAIR--DAQVNGEGQRPYMTDAE--RV 104  
DB 35 SAGAG-FARELGSRRSLDQKNTEDLSRPLKPPADSRALGEGWAKSKQLNDEDEKQ 93  
QY 105 DQARENGFNIVYDSIINSLSPDIGHNCSKRY-LETPNTSIIFPNBGMSSILR 163  
DB 94 BELIERAINIYLSDRISLHRHIEDKMYECKSQKFNRTLPPTSVAIAFNEAMSTLR 153  
QY 164 TVHSVLNRPEPELVAEIVLVDFSDREHLKPLEDYMALFPSPVRLRTKKEGHLRTML 223  
DB 154 TIHVLTSRAVLKEIILVDDLSDRVYLKTQLETYSINLDRVLLIRNKEGELVARLI 213  
QY 224 GASVATGDVITFLDSHCEANVNLPLLDRIARNRKTIVCPMIDVIDHDFRYETQAGDA 283  
DB 214 GATPATDVLTLFYLCHCEGNSGMLPELIERIGRYETAIVACVIDITDNTPEFYMQGEP 273  
QY 284 MRGAFDEMYKRIPIPEL--QKADPSDPESPVMAGLFAVDRKMWELGVDPELEI 341  
DB 274 MGGFDWRLTFQMSVPRQERDRISRIDPISPTMAGGLFAVSKKTFQYLGTYDTGAEV 333  
QY 342 WGSQVEYISFKVMCGGSMEDIPCSRVGHIYKTYVPYVPAVGLSARLKLVAEVMDEX 401  
DB 334 WGGENLISPRVWOCGKLEIHPCSHGIVHPKAPARP--NPLQNTAALAAVMDXY 390  
QY 402 AEYIYQRRPEYRHSAGDVAVQKLRSLNCKSRKFWMTKIAMDLPKFPYVBPAAAM- 460  
DB 391 KEHFYNNRPPARKAYGDISERKLLRELRCKSPDWYLNKVN--FPMILHVEDBRP--GWH 445  
QY 461 GEIRNVG-TGLCAD---TKHAGLSPLRLBGCVRGREAAMNNOVFTFWREDIRPDP 516  
DB 446 GAISRSGISSECLDYNPDNPTGANLSLFGC-HGQG---GNQFFEYTSNKEIRFNS- 498  
QY 517 QHTKKFCFDAISHTSPVTLVYDCHSMKNQMLKRYKDK-TLHVPVSGSCMDC---SES 569  
DB 499 --VTELCAEVEPQKRYVGMQKCPKDGFPVPAIIIMHFPREDCTIFHPHSGCLSLAVRPREG 556

QY 570 DRIEMNTCNBSLTQOMLF 590  
DB 557 RPDQKRTCDALDKNQIWSFE 577

RESULT 15  
US-09-217-306B-10  
Sequence 10, Application US/09217306B  
Patent No. 6465220  
GENERAL INFORMATION:  
APPLICANT: Haasan, Helle  
APPLICANT: Clausen, Henrik  
TITLE OF INVENTION: Glycosylation Using Galnac-T4 Transferase  
FILE REFERENCE: 8850\*1  
CURRENT APPLICATION NUMBER: US/09/217,306B  
CURRENT FILING DATE: 1998-12-21  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 10  
LENGTH: 560  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Interferon fragment-galnac-T4 predicted coding region  
NAME/KEY: PEPTIDE  
LOCATION: (1)..(25)  
OTHER INFORMATION: human interferon signal peptide  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (37)..(560)  
OTHER INFORMATION: Galnac-T4  
US-09-217-306B-10

Query Match 31.2%; Score 1024; DB 4; Length 560;  
Best Local Similarity 41.0%; Pred. No. 6.1e-101;  
Matches 213; Conservative 88; Mismatches 183; Indels 36; Gaps 14;

QY 90 GEOGRPYM-----TDARVDOARENGFNIVYDSIINSLSPDIGHNCSKRY-LETL 144  
DB 57 GEMGKAKKLQNLNDEDEKQBELIERAINIYLSDRISLHRHIEDKMYECKSQKFNRTLP 116  
QY 145 PNTSIIFPNBGMSSILRTVHSVLNRSPEPELVAEIVLVDFSDREHLKPLEDYMALFP 204  
DB 117 PPTSVAIAFNEAMSTLRLTIHSLTSPVALLKEIILVDDLSDRVYLKTQLETYSINL 176  
QY 205 SVRLRTKKEGHLRTMLGASVATGDVITFLDSHCEANVNLPLLDRIARNRKTIVCP 264  
DB 177 RVRLIRNKEGELVARLIGATPATDVLTLFYLCHCEGNSGMLPELIERIGRYETAIVCP 236  
QY 265 MIDVIDHDFRYETQAGDAAGAFDEMYKRIPIPEL--QKADPSDPESPVMAGGLF 322  
DB 237 VIDITDNTPEFYMQGEPMDRLTFQMSVPRQERDRISRIDPISPTMAGGLF 296  
QY 323 ADRKMFELGVDPELEINGEYETISFKVMCGGSMEDIPCSRVGHIYKTYVPYVPA 382  
DB 297 AVSKKTFQYLGTYDTGAEVWGGENLISLFRWOCGKLEIHPCSHGIVHPKAPARP- 355  
QY 383 GVSIAERLKVABWMDYAEYIYQRRPEYRHSAGDVAVQKLRSLNCKSRKFWMTKI 442  
DB 356 --NPLQNTAALAAVMDXYEHFNNRPPARKAYGDISERKLLRELRCKSPDWYLNKVN 413  
QY 443 AMDLPKFPYVBPAAAM-GEIRNVG-TGLCAD---TKHAGLSPLRLBGCVRGREAAM 497  
DB 414 --FPMILHVEDBRP--GMHGAISRSGISSECLDYNPDNPTGANLSLFGC-HGQG--- 463  
QY 498 MNNOVFTFWREDIRPDPQHTKKFCFDAISHTSPVTLVYDCHSMKNQMLKRYKDK 553  
DB 464 --GNQFFEYTSNKEIRFNS--VTELCAEVEPQKRYVGMQKCPKDGFPVPAIIIMHFPREDG 519  
QY 554 TVHVPVSGSCMDC---SESDBRI-FMNTCNBSLTQOMLF 590

Db 520 TIFPHSGTCLSAVTPTEGPRPDVQRTCDALDKQNIWFE 559

## RESULT 16

US-09-795-926-33

Sequence 33, Application US/09795926

Patent No. 6555669

GENERAL INFORMATION:

APPLICANT: Donoho, Gregory

APPLICANT: Hilbun, Erin

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Friedrich, Glenn

APPLICANT: Abuin, Alejandro

APPLICANT: Zambrowicz, Brian

APPLICANT: Sands, Arthur T.

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Hu, Yi

APPLICANT: Kieke, James Alvin

APPLICANT: Potter, David George

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795,926

PRIOR APPLICATION NUMBER: US 60/185,920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186,558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191,849

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 33

LENGTH: 240

TYPE: PRT

ORGANISM: homo sapiens

US-09-795-926-33

Query Match 30.2%; Score 990; DB 4; Length 240;  
Best Local Similarity 97.9%; Pred. No. 6,8e-98;  
Matches 189; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRREKRLQVAVLALVLLPVNGMALYREROPDTPGSGAAYAPAAQSGSHSROK 60

Db 1 MRREKRLQVAVLALVLLPVNGMALYREROPDTPGSGAAYAPAAQSGSHSROK 60

QY 61 KTFPLGQGLKDMHDKAIRDRAQVNGEGRPYPMTDARVDQAYRENGFNIVYSDK 120

Db 61 KTFPLGQGLKDMHDKAIRDRAQVNGEGRPYPMTDARVDQAYRENGFNIVYSDK 120

QY 121 ISLRSLEPDIRHPNCSKRYLETLPNTSIIIPFNHSGSSILRTVHSLNRSPPELVAEI 180

Db 121 ISLRSLEPDIRHPNCSKRYLETLPNTSIIIPFNHSGSSILRTVHSLNRSPPELVAEI 180

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

QY 181 VLVDPSDRHLK 193

Db 181 VLVDPSDRGRIR 193

QY 181 VLVDPSDRHLK 193

APPLICANT: Walke, D. Wade

APPLICANT: Wilganowski, Nathaniel L.

APPLICANT: Hu, Yi

APPLICANT: Kieke, James Alvin

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME

FILE REFERENCE: LEX-0144-USA

CURRENT APPLICATION NUMBER: US/09/795,926

PRIOR APPLICATION NUMBER: US 60/185,920

PRIOR FILING DATE: 2000-02-29

PRIOR APPLICATION NUMBER: US 60/186,558

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 60/191,849

PRIOR FILING DATE: 2000-03-24

NUMBER OF SEQ ID NOS: 47

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 37

LENGTH: 209

TYPE: PRT

ORGANISM: homo sapiens

US-09-795-926-37

Query Match 30.1%; Score 988; DB 4; Length 209;  
Best Local Similarity 100.0%; Pred. No. 8,8e-98;  
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRREKRLQVAVLALVLLPVNGMALYREROPDTPGSGAAYAPAAQSGSHSROK 60

Db 1 MRREKRLQVAVLALVLLPVNGMALYREROPDTPGSGAAYAPAAQSGSHSROK 60

QY 61 KTFPLGQGLKDMHDKAIRDRAQVNGEGRPYPMTDARVDQAYRENGFNIVYSDK 120

Db 61 KTFPLGQGLKDMHDKAIRDRAQVNGEGRPYPMTDARVDQAYRENGFNIVYSDK 120

QY 121 ISLRSLEPDIRHPNCSKRYLETLPNTSIIIPFNHSGSSILRTVHSLNRSPPELVAEI 180

Db 121 ISLRSLEPDIRHPNCSKRYLETLPNTSIIIPFNHSGSSILRTVHSLNRSPPELVAEI 180

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189

QY 181 VLVDPSDR 189

Db 181 VLVDPSDR 189









Db 487 -----RAILYPCHGM-SGQLVRYXSADGLQLGLGSTAFLPDS-KCLVDDG 530  
QY 569 SDHRIFMNTCNPSLSLTOQWLFEHTNS 594  
Db 531 TGRMPTLKKECDVAPRTORLMDFTOS 556

## RESULT 25

US-09-270-767-45334  
; Sequence 45334, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45334  
; LENGTH: 289  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-45334

Query Match 20.6%; Score 675; DB 4; Length 289;

Best Local Similarity 50.2%; Pred. No. 8.2e-64; Matches 131; Conservative 45; Mismatches 77; Indels 8; Gaps 4;

QY 111 NGFNIVYSDKISLNRSPLPDIRHPNCNSKRYLETLPNTSIIIPHNEGWSSILRTVHSYLN 170  
Db 34 NRFNGEASDALPSNRDIPDTENPMCRCTKCYREDLPETSVITTFHNEARSTLLRTIVSYLN 93  
QY 171 RSPPELVAELIVLDDPSDREHLKKPLEDYMALFPESVRLRTKCKEGLIRTRMLGASVATG 230  
Db 94 RSPPELIRELIVLDDPSD--HPEDGLE--LAKIDKRVIRNDKREGLVRSRVKGAAYVS 149  
QY 231 DVITFLDSHCANVNWLPPLDRIARNRKTIVCPMIDVIDHDDPFRYETQAGDAMRGAPDW 290  
Db 150 SVLTFPLDSHVCENMWELEPLERVAEDPTRVVCPVIDVISMDNFOYIGASAD-LRGGFDM 208  
QY 291 EMYKRIPIPE--LQKADPSDPESPVMAGLPAVDKMKFWELAGYDPGLBIWGEQY 347  
Db 209 NLIFKWEYLSPSERAMRNDPTTAIRTEMIGLGLFVIDKAYFNKLKGYDMKMDVWGENTL 268  
QY 348 EISFKVMCGGRMEDIPCSR 368  
Db 269 EISFRVMQGGSLRIIPCSR 289

Search completed: November 22, 2004, 13:44:43  
Job time : 23 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 13:37:52 ; Search time 57 Seconds  
(Without alignments)  
3746.297 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278

Sequence: 1 MRKERRLLQAVLVLAALV.....TQQLPFRHTNSTYLEKFNRY 603

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1570615 seqs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppaa/PCR\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/PCRUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US09C\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubppaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*  
20: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3278	100.0	603	9 US-09-795-926-43	Sequence 43, App1
2	3278	100.0	603	13 US-10-001-851-2	Sequence 2, App1
3	3278	100.0	603	14 US-10-364-774-43	Sequence 43, App1
4	3242.5	98.9	631	9 US-09-795-926-41	Sequence 41, App1
5	3242.5	98.9	631	14 US-10-364-774-41	Sequence 41, App1
6	2771	84.5	506	9 US-09-795-926-31	Sequence 31, App1
7	2771	84.5	506	14 US-10-364-774-31	Sequence 31, App1
8	2746.5	83.8	535	9 US-09-795-926-29	Sequence 29, App1
9	2746.5	83.8	535	14 US-10-364-774-29	Sequence 29, App1
10	1877	57.3	366	9 US-09-795-926-39	Sequence 39, App1
11	1877	57.3	366	14 US-10-364-774-39	Sequence 39, App1
12	1392	42.5	276	14 US-10-292-896-3	Sequence 3, App1
13	1370	41.8	269	9 US-09-795-926-27	Sequence 27, App1

14	1370	41.8	269	14 US-10-364-774-27	Sequence 27, App1
15	1311	40.0	321	9 US-09-795-926-35	Sequence 35, App1
16	1311	40.0	321	14 US-10-364-774-35	Sequence 35, App1
17	1125	34.3	561	9 US-09-925-301-1006	Sequence 1006, App1
18	1117	34.1	559	13 US-10-001-851-24	Sequence 24, App1
19	1116	34.0	559	14 US-10-001-851-21	Sequence 21, App1
20	1116	34.0	559	14 US-10-205-219-36	Sequence 36, App1
21	1115	34.0	559	14 US-10-205-219-76	Sequence 76, App1
22	1115	34.0	559	13 US-10-001-851-22	Sequence 22, App1
23	1115	34.0	559	13 US-10-001-851-23	Sequence 23, App1
24	1113	33.7	556	14 US-10-292-896-62	Sequence 62, App1
25	1106	32.9	626	13 US-10-001-851-27	Sequence 27, App1
26	1079	32.0	578	14 US-10-292-896-118	Sequence 118, App1
27	1050	31.5	578	13 US-10-074-527-8	Sequence 8, App1
28	1033	31.4	657	14 US-10-341-434-32	Sequence 32, App1
29	1029	30.5	558	14 US-10-292-896-66	Sequence 66, App1
30	999.5	30.5	558	15 US-10-433-256-6	Sequence 6, App1
31	999.5	30.2	581	13 US-10-074-527-2	Sequence 2, App1
32	991.5	30.2	581	15 US-10-085-198-122	Sequence 122, App1
33	991.5	30.2	581	15 US-10-433-256-4	Sequence 4, App1
34	991.5	30.2	581	15 US-10-789-241-44	Sequence 44, App1
35	991.5	30.2	240	9 US-09-795-926-33	Sequence 33, App1
36	990	30.2	240	14 US-10-364-774-33	Sequence 33, App1
37	990	30.2	612	13 US-10-001-851-25	Sequence 25, App1
38	988.5	30.1	209	9 US-09-795-926-37	Sequence 37, App1
39	988	30.1	209	14 US-10-364-774-37	Sequence 37, App1
40	988	29.6	581	14 US-10-292-896-58	Sequence 58, App1
41	971.5	29.6	552	9 US-09-815-028-2	Sequence 2, App1
42	971	29.6	552	14 US-10-028-072-196	Sequence 196, App1
43	971	29.6	552	14 US-10-140-808-196	Sequence 196, App1
44	971	29.6	552	14 US-10-140-808-196	Sequence 196, App1
45	971	29.6	552	14 US-10-121-004-196	Sequence 196, App1
46	971	29.6	552	14 US-10-123-904-196	Sequence 196, App1
47	971	29.6	552	14 US-10-140-474-196	Sequence 196, App1
48	971	29.6	552	14 US-10-175-748-196	Sequence 196, App1
49	971	29.6	552	14 US-10-176-921-196	Sequence 196, App1
50	971	29.6	552	14 US-10-176-921-196	Sequence 196, App1
51	971	29.6	552	14 US-10-137-865-196	Sequence 196, App1
52	971	29.6	552	14 US-10-140-474-196	Sequence 196, App1
53	971	29.6	552	14 US-10-140-474-196	Sequence 196, App1
54	971	29.6	552	14 US-10-143-114-196	Sequence 196, App1
55	971	29.6	552	14 US-10-143-114-196	Sequence 196, App1
56	971	29.6	552	14 US-10-140-002-196	Sequence 196, App1
57	971	29.6	552	14 US-10-142-419-196	Sequence 196, App1
58	971	29.6	552	14 US-10-123-261-196	Sequence 196, App1
59	971	29.6	552	14 US-10-142-423-196	Sequence 196, App1
60	971	29.6	552	14 US-10-121-050-196	Sequence 196, App1
61	971	29.6	552	14 US-10-141-755-196	Sequence 196, App1
62	971	29.6	552	14 US-10-143-033-196	Sequence 196, App1
63	971	29.6	552	14 US-10-123-103-196	Sequence 196, App1
64	971	29.6	552	14 US-10-123-236-196	Sequence 196, App1
65	971	29.6	552	14 US-10-123-261-196	Sequence 196, App1
66	971	29.6	552	14 US-10-140-921-196	Sequence 196, App1
67	971	29.6	552	14 US-10-140-928-196	Sequence 196, App1
68	971	29.6	552	14 US-10-121-045-196	Sequence 196, App1
69	971	29.6	552	14 US-10-123-293-196	Sequence 196, App1
70	971	29.6	552	14 US-10-123-903-196	Sequence 196, App1
71	971	29.6	552	14 US-10-124-819-196	Sequence 196, App1
72	971	29.6	552	14 US-10-124-822-196	Sequence 196, App1
73	971	29.6	552	14 US-10-140-925-196	Sequence 196, App1
74	971	29.6	552	14 US-10-160-498-196	Sequence 196, App1
75	971	29.6	552	14 US-10-124-824-196	Sequence 196, App1
76	971	29.6	552	14 US-10-127-825A-196	Sequence 196, App1
77	971	29.6	552	14 US-10-127-825A-196	Sequence 196, App1
78	971	29.6	552	14 US-10-127-835A-196	Sequence 196, App1
79	971	29.6	552	14 US-10-127-835A-196	Sequence 196, App1
80	971	29.6	552	14 US-10-127-901A-196	Sequence 196, App1
81	971	29.6	552	14 US-10-128-693A-196	Sequence 196, App1
82	971	29.6	552	14 US-10-131-818A-196	Sequence 196, App1
83	971	29.6	552	14 US-10-131-818A-196	Sequence 196, App1
84	971	29.6	552	14 US-10-131-824A-196	Sequence 196, App1
85	971	29.6	552	14 US-10-131-824A-196	Sequence 196, App1
86	971	29.6	552	14 US-10-131-830A-196	Sequence 196, App1

87 971 29.6 552 14 US-10-137-872A-196 Sequence 196, App  
88 971 29.6 552 14 US-10-147-500-196 Sequence 196, App  
89 971 29.6 552 14 US-10-147-502-196 Sequence 196, App  
90 971 29.6 552 14 US-10-147-515-196 Sequence 196, App  
91 971 29.6 552 14 US-10-147-517-196 Sequence 196, App  
92 971 29.6 552 14 US-10-147-526-196 Sequence 196, App  
93 971 29.6 552 14 US-10-147-527-196 Sequence 196, App  
94 971 29.6 552 14 US-10-121-041-196 Sequence 196, App  
95 971 29.6 552 14 US-10-121-043-196 Sequence 196, App  
96 971 29.6 552 14 US-10-121-047-196 Sequence 196, App  
97 971 29.6 552 14 US-10-123-215-196 Sequence 196, App  
98 971 29.6 552 14 US-10-123-902-196 Sequence 196, App  
99 971 29.6 552 14 US-10-123-908-196 Sequence 196, App  
100 971 29.6 552 14 US-10-123-909-196 Sequence 196, App

## ALIGNMENTS

RESULT 1  
US-09-795-926-43

Sequence 43, Application US/09795926  
Patent No. US20020098486A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Halbur, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedlich, Glenn  
APPLICANT: Abulin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milgnowski, Nathaniel L.  
APPLICANT: Kleke, James Alvin  
APPLICANT: Potler, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: LEX-0144-USA  
CURRENT APPLICATION NUMBER: US/09/795, 926  
CURRENT FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/185, 920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186, 558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191, 849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 43  
LENGTH: 603  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-43

Query Match 100.0%; Score 3278; DB 9; Length 603;  
Best Local Similarity 100.0%; Pred. No. 5e-310;  
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRERKLLQAVLALVLAALVLLPNVGLMALYRERQPDGTGGSGAAVAAPAAQSGSHSRQK 60  
DB 1 MRRERKLLQAVLALVLAALVLLPNVGLMALYRERQPDGTGGSGAAVAAPAAQSGSHSRQK 60  
QY 61 KTFPLGSGQKLDKMDHDEKAIIRDARVNGEQGRPYMTDAERVDQAARENGFNIIYVSDK 120  
DB 61 KTFPLGSGQKLDKMDHDEKAIIRDARVNGEQGRPYMTDAERVDQAARENGFNIIYVSDK 120  
QY 121 ISLRSPLPDIRHPNCNSKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVNSPPELVAEI 180  
DB 121 ISLRSPLPDIRHPNCNSKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVNSPPELVAEI 180  
QY 181 VLVDPSDRHLKKPLLEDYMAIPPSVRIILRTKREGLIRTRMLGASVATGDIVITFLDSHC 240  
DB 181 VLVDPSDRHLKKPLLEDYMAIPPSVRIILRTKREGLIRTRMLGASVATGDIVITFLDSHC 240

DB 181 VLVDPSDRHLKKPLLEDYMAIPPSVRIILRTKREGLIRTRMLGASVATGDIVITFLDSHC 240  
QY 241 EAVNWMLPPLDIRIARNRKTIYCPMTDIVIDHDDPERYETQADARGAFFDMEMYYKRIPIR 300  
DB 241 EAVNWMLPPLDIRIARNRKTIYCPMTDIVIDHDDPERYETQADARGAFFDMEMYYKRIPIR 300  
QY 301 PELQKADPSDFPESPVNAGGLFAVDRKMFWEELGGYDGLBIWGGEOYEISFKVMCCGRM 360  
DB 301 PELQKADPSDFPESPVNAGGLFAVDRKMFWEELGGYDGLBIWGGEOYEISFKVMCCGRM 360  
QY 361 EDIPCSRVGHIIYKTYVYKYPAGVSLARNIKRAVAEMVMDYAEIYRREBRYRLSGDV 420  
DB 361 EDIPCSRVGHIIYKTYVYKYPAGVSLARNIKRAVAEMVMDYAEIYRREBRYRLSGDV 420  
QY 421 AVQKGLSSLNCSFKFMFTKIAMDLPKFPVPEPPAAANGELRNVTGICADTKHAGLG 480  
DB 421 AVQKGLSSLNCSFKFMFTKIAMDLPKFPVPEPPAAANGELRNVTGICADTKHAGLG 480  
QY 481 SPLRLBEGCVARGGEAANNMNVFTFTWRREDIRPGDPQHTKKFCDAISHTSPVTLYDCHS 540  
DB 481 SPLRLBEGCVARGGEAANNMNVFTFTWRREDIRPGDPQHTKKFCDAISHTSPVTLYDCHS 540  
QY 541 MKGNQLMKTRKDKTLVHPVSGSCMDCSBSDRIFPANTCNSSLTQOWLFEHTNSTYLEKF 600  
DB 541 MKGNQLMKTRKDKTLVHPVSGSCMDCSBSDRIFPANTCNSSLTQOWLFEHTNSTYLEKF 600  
QY 601 NRN 603  
DB 601 NRN 603

## RESULT 2

US-10-001-851-2  
Sequence 2, Application US/10001851  
Publication No. US20020115628A1  
GENERAL INFORMATION:  
APPLICANT: MEYERS, Rachel A.  
APPLICANT: WILLIMSON, Mark  
TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1 Human Glycosyl Transferase  
FILE REFERENCE: 10147-5601  
CURRENT APPLICATION NUMBER: US/10/001, 851  
CURRENT FILING DATE: 2001-11-20  
PRIOR APPLICATION NUMBER: US 60/249, 939  
PRIOR FILING DATE: 2000-11-20  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: Patencin Ver. 2.1  
SEQ ID NO 2  
LENGTH: 603  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-001-851-2

Query Match 100.0%; Score 3278; DB 13; Length 603;  
Best Local Similarity 100.0%; Pred. No. 5e-310;  
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRERKLLQAVLALVLAALVLLPNVGLMALYRERQPDGTGGSGAAVAAPAAQSGSHSRQK 60  
DB 1 MRRERKLLQAVLALVLAALVLLPNVGLMALYRERQPDGTGGSGAAVAAPAAQSGSHSRQK 60  
QY 61 KTFPLGSGQKLDKMDHDEKAIIRDARVNGEQGRPYMTDAERVDQAARENGFNIIYVSDK 120  
DB 61 KTFPLGSGQKLDKMDHDEKAIIRDARVNGEQGRPYMTDAERVDQAARENGFNIIYVSDK 120  
QY 121 ISLRSPLPDIRHPNCNSKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVNSPPELVAEI 180  
DB 121 ISLRSPLPDIRHPNCNSKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVNSPPELVAEI 180  
QY 181 VLVDPSDRHLKKPLLEDYMAIPPSVRIILRTKREGLIRTRMLGASVATGDIVITFLDSHC 240  
DB 181 VLVDPSDRHLKKPLLEDYMAIPPSVRIILRTKREGLIRTRMLGASVATGDIVITFLDSHC 240

QY 241 EAVNMLPPLDRIARNRKTIVCPMIDVIDHDDEFRYTOAGDMRGAFDMEYKRIPI 300  
DB 241 EAVNMLPPLDRIARNRKTIVCPMIDVIDHDDEFRYTOAGDMRGAFDMEYKRIPI 300  
QY 301 PELQADPSDFPESPVNAGGLFAVDRKMFELGSDYDGLIWSGEQYEISFKVMCCGRM 360  
DB 301 PELQADPSDFPESPVNAGGLFAVDRKMFELGSDYDGLIWSGEQYEISFKVMCCGRM 360  
QY 361 EDIPCSRVGHIYRKVYKYPAGVSLARNLKRVAEVMMDYAEIYORREYRHLSDGV 420  
DB 361 EDIPCSRVGHIYRKVYKYPAGVSLARNLKRVAEVMMDYAEIYORREYRHLSDGV 420  
QY 421 AVQKLRSSLNCKSFKFMFKTIAMDLPKFPYPPVPPAAANGELRNVTGICADTKHGALG 480  
DB 421 AVQKLRSSLNCKSFKFMFKTIAMDLPKFPYPPVPPAAANGELRNVTGICADTKHGALG 480  
QY 481 SPLRLEGCVGRGEAANNMNOVFTFTWRREDIRPGDPQHTKKFCFDALSHTSPTLYDCHS 540  
DB 481 SPLRLEGCVGRGEAANNMNOVFTFTWRREDIRPGDPQHTKKFCFDALSHTSPTLYDCHS 540  
QY 541 MKGNQLMKRYKDKTLVHPVSGSCMDCSESDHRIFMNTCNPSLTOQWLFHTNSTVLEKF 600  
DB 541 MKGNQLMKRYKDKTLVHPVSGSCMDCSESDHRIFMNTCNPSLTOQWLFHTNSTVLEKF 600  
QY 601 NRN 603  
DB 601 NRN 603

## RESULT 3

US-10-364-774-43  
; Sequence 43, Application US/10364774  
; Publication No. US20030144497A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedlich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sander, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kiege, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/10/364,774  
; CURRENT FILING DATE: 2003-02-11  
; PRIOR APPLICATION NUMBER: US/09/795,926  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 43  
; LENGTH: 603  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-364-774-43

Query Match 100.0%; Score 3278; DB 14; Length 603;  
Best Local Similarity 100.0%; Pctd. No. 5e-310;  
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MRRERKLLQAVLVLAALVTLVNVGLMALYRERQDPGTGCGSAVAAPAGQSHSRQK 60  
DB 1 MRRERKLLQAVLVLAALVTLVNVGLMALYRERQDPGTGCGSAVAAPAGQSHSRQK 60

DB 1 MRRERKLLQAVLVLAALVTLVNVGLMALYRERQDPGTGCGSAVAAPAGQSHSRQK 60  
QY 61 KTFPLGSGOGLKQWHDDEALIRDAORVNGEOGRPYMTAERDQYRENGFPIYVSDK 120  
DB 61 KTFPLGSGOGLKQWHDDEALIRDAORVNGEOGRPYMTAERDQYRENGFPIYVSDK 120  
QY 121 ISLNRSLPDRHPNCKSRVLETLPTNTSIIIPFNESGSLRTVHVLNRSPELVAEI 180  
DB 121 ISLNRSLPDRHPNCKSRVLETLPTNTSIIIPFNESGSLRTVHVLNRSPELVAEI 180  
QY 181 VLVDVDFSRHLLKKPLEDYNALPPSVAILRTKREGILIRTMIGASVATGDVITFLDSHC 240  
DB 181 VLVDVDFSRHLLKKPLEDYNALPPSVAILRTKREGILIRTMIGASVATGDVITFLDSHC 240  
QY 241 EAVNMLPPLDRIARNRKTIVCPMIDVIDHDDEFRYTOAGDMRGAFDMEYKRIPI 300  
DB 241 EAVNMLPPLDRIARNRKTIVCPMIDVIDHDDEFRYTOAGDMRGAFDMEYKRIPI 300  
QY 301 PELQADPSDFPESPVNAGGLFAVDRKMFELGSDYDGLIWSGEQYEISFKVMCCGRM 360  
DB 301 PELQADPSDFPESPVNAGGLFAVDRKMFELGSDYDGLIWSGEQYEISFKVMCCGRM 360  
QY 361 EDIPCSRVGHIYRKVYKYPAGVSLARNLKRVAEVMMDYAEIYORREYRHLSDGV 420  
DB 361 EDIPCSRVGHIYRKVYKYPAGVSLARNLKRVAEVMMDYAEIYORREYRHLSDGV 420  
QY 421 AVQKLRSSLNCKSFKFMFKTIAMDLPKFPYPPVPPAAANGELRNVTGICADTKHGALG 480  
DB 421 AVQKLRSSLNCKSFKFMFKTIAMDLPKFPYPPVPPAAANGELRNVTGICADTKHGALG 480  
QY 481 SPLRLEGCVGRGEAANNMNOVFTFTWRREDIRPGDPQHTKKFCFDALSHTSPTLYDCHS 540  
DB 481 SPLRLEGCVGRGEAANNMNOVFTFTWRREDIRPGDPQHTKKFCFDALSHTSPTLYDCHS 540  
QY 541 MKGNQLMKRYKDKTLVHPVSGSCMDCSESDHRIFMNTCNPSLTOQWLFHTNSTVLEKF 600  
DB 541 MKGNQLMKRYKDKTLVHPVSGSCMDCSESDHRIFMNTCNPSLTOQWLFHTNSTVLEKF 600  
QY 601 NRN 603  
DB 601 NRN 603

## RESULT 4

US-09-795-926-41  
; Sequence 41, Application US/09795926  
; Patent No. US20020098486A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedlich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sander, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kiege, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/09/795,926  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 41  
LENGTH: 631  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-41

Query Match 98.9%; Score 3242.5; DB 9; Length 631;  
Best Local Similarity 95.4%; Pred. No. 1.5e-306;  
Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

```
QY 1 MRRREKLLQAVLVLAALVLLPVGGLMALYRERQPDGTGSGAAVAAPAGGSHSRQK 60
DB 1 MRRREKLLQAVLVLAALVLLPVGGLMALYRERQPDGTGSGAAVAAPAGGSHSRQK 60
QY 61 KTFPLGGOQKLDKMDHDEAIRDARVNGEGRPYPTDAERVDQAYRENGFNIVSDK 120
DB 61 KTFPLGGOQKLDKMDHDEAIRDARVNGEGRPYPTDAERVDQAYRENGFNIVSDK 120
QY 121 ISLRSPLDIHHPNCNKRILETLPNTSIIIPFNHEGMSLLRTVHSLNRSPELVAEI 180
DB 121 ISLRSPLDIHHPNCNKRILETLPNTSIIIPFNHEGMSLLRTVHSLNRSPELVAEI 180
QY 181 VLVDPSDRHLKKPLDYMALFPSVRLRTKKREGILRTMLGASVATGDIVITFLDSC 240
DB 181 VLVDPSDRHLKKPLDYMALFPSVRLRTKKREGILRTMLGASVATGDIVITFLDSC 240
QY 241 EANNMPLPLDRIARNRKTIIVCPMIDVIDHDDFRYETQADAMRGAFDMEYKRIPI 300
DB 241 EANNMPLPLDRIARNRKTIIVCPMIDVIDHDDFRYETQADAMRGAFDMEYKRIPI 300
QY 301 PELQKADPSDFESPVMAGSLFAVDRKMFELGSDYDGLIHWGEQYEISFKGLHMLPRL 360
DB 301 PELQKADPSDFESPVMAGSLFAVDRKMFELGSDYDGLIHWGEQYEISFKGLHMLPRL 360
QY 353 -----VMMCGRMEDIPCSRGHYIRKYVPYKVPAGVSLAARNLK 391
DB 353 -----VMMCGRMEDIPCSRGHYIRKYVPYKVPAGVSLAARNLK 391
QY 392 RVAEVMNDEVAEYIYQRRPEYRHLASGDVAVQKLRSLNCKSFKMTKIAMLPKFP 451
DB 392 RVAEVMNDEVAEYIYQRRPEYRHLASGDVAVQKLRSLNCKSFKMTKIAMLPKFP 451
QY 421 RVAEVMNDEVAEYIYQRRPEYRHLASGDVAVQKLRSLNCKSFKMTKIAMLPKFP 480
DB 421 RVAEVMNDEVAEYIYQRRPEYRHLASGDVAVQKLRSLNCKSFKMTKIAMLPKFP 480
QY 452 PVEPPAAAMGEIRNVGTGLCADTGHGALGSLRLGCVRGGEAANNMQUFTFTWREDI 511
DB 452 PVEPPAAAMGEIRNVGTGLCADTGHGALGSLRLGCVRGGEAANNMQUFTFTWREDI 511
QY 512 RPDGPQHTKKCFDAISHTSPVTLTDCHSMKGNQIMKRYKDKTLVHPVSGSCMDCSESDH 571
DB 512 RPDGPQHTKKCFDAISHTSPVTLTDCHSMKGNQIMKRYKDKTLVHPVSGSCMDCSESDH 571
QY 541 RPDGPQHTKKCFDAISHTSPVTLTDCHSMKGNQIMKRYKDKTLVHPVSGSCMDCSESDH 600
DB 541 RPDGPQHTKKCFDAISHTSPVTLTDCHSMKGNQIMKRYKDKTLVHPVSGSCMDCSESDH 600
QY 572 RIFMNTCNPSLSLTQOMLFEHTNSTVLEKEN 601
DB 572 RIFMNTCNPSLSLTQOMLFEHTNSTVLEKEN 601
QY 601 RIFMNTCNPSLSLTQOMLFEHTNSTVLEKEN 630
DB 601 RIFMNTCNPSLSLTQOMLFEHTNSTVLEKEN 630
```

## RESULT 5

US-10-364-774-41  
Sequence 41, Application US/10364774  
Publication No. US20030144497A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedlich, Glenn  
APPLICANT: Abutin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kieke, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
FILE REFERENCE: LEX-0144-USA  
CURRENT APPLICATION NUMBER: US/10/364, 774  
CURRENT FILING DATE: 2003-02-11  
PRIOR APPLICATION NUMBER: US/09/795, 926  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/185, 920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186, 558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191, 849  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 41  
LENGTH: 631  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-364-774-41

Query Match 98.9%; Score 3242.5; DB 14; Length 631;  
Best Local Similarity 95.4%; Pred. No. 1.5e-306;  
Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

```
QY 1 MRRREKLLQAVLVLAALVLLPVGGLMALYRERQPDGTGSGAAVAAPAGGSHSRQK 60
DB 1 MRRREKLLQAVLVLAALVLLPVGGLMALYRERQPDGTGSGAAVAAPAGGSHSRQK 60
QY 61 KTFPLGGOQKLDKMDHDEAIRDARVNGEGRPYPTDAERVDQAYRENGFNIVSDK 120
DB 61 KTFPLGGOQKLDKMDHDEAIRDARVNGEGRPYPTDAERVDQAYRENGFNIVSDK 120
QY 121 ISLRSPLDIHHPNCNKRILETLPNTSIIIPFNHEGMSLLRTVHSLNRSPELVAEI 180
DB 121 ISLRSPLDIHHPNCNKRILETLPNTSIIIPFNHEGMSLLRTVHSLNRSPELVAEI 180
QY 181 VLVDPSDRHLKKPLDYMALFPSVRLRTKKREGILRTMLGASVATGDIVITFLDSC 240
DB 181 VLVDPSDRHLKKPLDYMALFPSVRLRTKKREGILRTMLGASVATGDIVITFLDSC 240
QY 241 EANNMPLPLDRIARNRKTIIVCPMIDVIDHDDFRYETQADAMRGAFDMEYKRIPI 300
DB 241 EANNMPLPLDRIARNRKTIIVCPMIDVIDHDDFRYETQADAMRGAFDMEYKRIPI 300
QY 301 PELQKADPSDFESPVMAGSLFAVDRKMFELGSDYDGLIHWGEQYEISFKGLHMLPRL 360
DB 301 PELQKADPSDFESPVMAGSLFAVDRKMFELGSDYDGLIHWGEQYEISFKGLHMLPRL 360
QY 353 -----VMMCGRMEDIPCSRGHYIRKYVPYKVPAGVSLAARNLK 391
DB 353 -----VMMCGRMEDIPCSRGHYIRKYVPYKVPAGVSLAARNLK 391
QY 392 RVAEVMNDEVAEYIYQRRPEYRHLASGDVAVQKLRSLNCKSFKMTKIAMLPKFP 451
DB 392 RVAEVMNDEVAEYIYQRRPEYRHLASGDVAVQKLRSLNCKSFKMTKIAMLPKFP 451
QY 421 RVAEVMNDEVAEYIYQRRPEYRHLASGDVAVQKLRSLNCKSFKMTKIAMLPKFP 480
DB 421 RVAEVMNDEVAEYIYQRRPEYRHLASGDVAVQKLRSLNCKSFKMTKIAMLPKFP 480
QY 452 PVEPPAAAMGEIRNVGTGLCADTGHGALGSLRLGCVRGGEAANNMQUFTFTWREDI 511
DB 452 PVEPPAAAMGEIRNVGTGLCADTGHGALGSLRLGCVRGGEAANNMQUFTFTWREDI 511
QY 481 PVEPPAAAMGEIRNVGTGLCADTGHGALGSLRLGCVRGGEAANNMQUFTFTWREDI 540
DB 481 PVEPPAAAMGEIRNVGTGLCADTGHGALGSLRLGCVRGGEAANNMQUFTFTWREDI 540
QY 512 RPDGPQHTKKCFDAISHTSPVTLTDCHSMKGNQIMKRYKDKTLVHPVSGSCMDCSESDH 571
DB 512 RPDGPQHTKKCFDAISHTSPVTLTDCHSMKGNQIMKRYKDKTLVHPVSGSCMDCSESDH 571
QY 541 RPDGPQHTKKCFDAISHTSPVTLTDCHSMKGNQIMKRYKDKTLVHPVSGSCMDCSESDH 600
DB 541 RPDGPQHTKKCFDAISHTSPVTLTDCHSMKGNQIMKRYKDKTLVHPVSGSCMDCSESDH 600
QY 572 RIFMNTCNPSLSLTQOMLFEHTNSTVLEKEN 601
DB 572 RIFMNTCNPSLSLTQOMLFEHTNSTVLEKEN 601
QY 601 RIFMNTCNPSLSLTQOMLFEHTNSTVLEKEN 630
DB 601 RIFMNTCNPSLSLTQOMLFEHTNSTVLEKEN 630
```

## RESULT 6

US-09-795-926-31  
Sequence 31, Application US/09795926  
Patent No. US2002009486A1



Qy 578 CNPSSLTQOMLFEHTNSTVLEKFNRN 603  
Db 481 CNPSSLTQOMLFEHTNSTVLEKFNRN 506

RESULT 8  
US-09-795-926-29

Sequence 29, Application US/09795926  
Patent No. US20020098486A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abulin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kleke, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: LEX-0144-USA  
CURRENT FILING DATE: 2001-02-28  
PRIOR FILING DATE: 2000-02-29  
PRIOR FILING DATE: 2000-02-29  
PRIOR FILING DATE: 2000-02-29  
PRIOR FILING DATE: 2000-03-02  
PRIOR FILING DATE: 2000-03-02  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 29  
LENGTH: 535  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-29

Query Match 83.8%; Score 2746.5; DB 9; Length 535;  
Best Local Similarity 94.6%; Pred. No. 2.8e-258;  
Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

Qy 98 MTDARVDQARENGFNIVYSDKISLRSLPDIRHPNCSKRYLETLPNTSIIIPFNNEG 157  
Db 1 MTDARVDQARENGFNIVYSDKISLRSLPDIRHPNCSKRYLETLPNTSIIIPFNNEG 60  
Qy 158 WSSLIRTVHSLVNSPELVAEIVLVDDPSDREHLKKPLEDYMALPFSVRILRTKKEGL 217  
Db 61 WSSLIRTVHSLVNSPELVAEIVLVDDPSDREHLKKPLEDYMALPFSVRILRTKKEGL 120  
Qy 218 IRTMLGASVATGVTITFLDSHCANVWMLPPLDIRIARNKRTIVCPMIDIVIDHDDFRYE 277  
Db 121 IRTMLGASVATGVTITFLDSHCANVWMLPPLDIRIARNKRTIVCPMIDIVIDHDDFRYE 180  
Qy 278 TQAGDARGAFDWMYKRIPIPELOKADPSDFESPVAAGLFAVDRKWFWEIGGYDP 337  
Db 181 TQAGDARGAFDWMYKRIPIPELOKADPSDFESPVAAGLFAVDRKWFWEIGGYDP 240  
Qy 338 GLEIWGSGQVEISFK-----VMMGGRMEDIPCSRV 368  
Db 241 GLEIWGSGQVEISFKGLHMLPRLVSNWPOAVFLPRAPNMLALOVMMCGRMEDIPCSRV 300  
Qy 369 GHIRYKTVYKVPAGVSLARNLKRVAEVMDEVAEYIYORRPEYRHLISAGDAVQKRLS 428  
Db 301 GHIRYKTVYKVPAGVSLARNLKRVAEVMDEVAEYIYORRPEYRHLISAGDAVQKRLS 360  
Qy 429 SLNCKSFQWNTKIAMDLPKFYPPVEPPAAWGEIRNVTGLCADTKHAGLSPLRLEG 488  
Db 361 SLNCKSFQWNTKIAMDLPKFYPPVEPPAAWGEIRNVTGLCADTKHAGLSPLRLEG 420

Qy 489 VRGRGAAMNNMQVFTETWREDIRPGDPQHTKKECFDAISHTSPVTLYDCHSMKGNLWK 548  
Db 421 VRGRGAAMNNMQVFTETWREDIRPGDPQHTKKECFDAISHTSPVTLYDCHSMKGNLWK 480  
Qy 549 YRKDKTLYHPVSGSCMDCSSDRHIFMNTCNPSSLTQOMLFEHTNSTVLEKFNRN 603  
Db 481 YRKDKTLYHPVSGSCMDCSSDRHIFMNTCNPSSLTQOMLFEHTNSTVLEKFNRN 535

RESULT 9  
US-10-364-774-29

Sequence 29, Application US/10364774  
Publication No. US20030144497A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedrich, Glenn  
APPLICANT: Abulin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Milganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kleke, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
FILE REFERENCE: LEX-0144-USA  
CURRENT FILING DATE: 2003-02-11  
PRIOR FILING DATE: 2001-02-28  
PRIOR FILING DATE: 2001-02-28  
PRIOR FILING DATE: 2000-02-29  
PRIOR FILING DATE: 2000-02-29  
PRIOR FILING DATE: 2000-03-02  
PRIOR FILING DATE: 2000-03-02  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 29  
LENGTH: 535  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-364-774-29

Query Match 83.8%; Score 2746.5; DB 14; Length 535;  
Best Local Similarity 94.6%; Pred. No. 2.8e-258;  
Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

Qy 98 MTDARVDQARENGFNIVYSDKISLRSLPDIRHPNCSKRYLETLPNTSIIIPFNNEG 157  
Db 1 MTDARVDQARENGFNIVYSDKISLRSLPDIRHPNCSKRYLETLPNTSIIIPFNNEG 60  
Qy 158 WSSLIRTVHSLVNSPELVAEIVLVDDPSDREHLKKPLEDYMALPFSVRILRTKKEGL 217  
Db 61 WSSLIRTVHSLVNSPELVAEIVLVDDPSDREHLKKPLEDYMALPFSVRILRTKKEGL 120  
Qy 218 IRTMLGASVATGVTITFLDSHCANVWMLPPLDIRIARNKRTIVCPMIDIVIDHDDFRYE 277  
Db 121 IRTMLGASVATGVTITFLDSHCANVWMLPPLDIRIARNKRTIVCPMIDIVIDHDDFRYE 180  
Qy 278 TQAGDARGAFDWMYKRIPIPELOKADPSDFESPVAAGLFAVDRKWFWEIGGYDP 337  
Db 181 TQAGDARGAFDWMYKRIPIPELOKADPSDFESPVAAGLFAVDRKWFWEIGGYDP 240  
Qy 338 GLEIWGSGQVEISFK-----VMMGGRMEDIPCSRV 368  
Db 241 GLEIWGSGQVEISFKGLHMLPRLVSNWPOAVFLPRAPNMLALOVMMCGRMEDIPCSRV 300  
Qy 369 GHIRYKTVYKVPAGVSLARNLKRVAEVMDEVAEYIYORRPEYRHLISAGDAVQKRLS 428

Db 301 GHYRKVPPKVPAGVSLARLKVAVMDEVAEYIYORRPERHLSAGDVAVQKLR 360  
Qy 429 SLNCKSFEMTKIAMDLPKPYPEVPEPPAAWGEIRNVGTGLCADYTKHAGLSPLREGC 488  
Db 361 SLNCKSFEMTKIAMDLPKPYPEVPEPPAAWGEIRNVGTGLCADYTKHAGLSPLREGC 420  
Qy 489 VRRGGEAANNMNVFTTWEDIRPGDPQHTKKCFPAISHTSPVTLYDCSHKMGNDLWK 548  
Db 421 VRRGGEAANNMNVFTTWEDIRPGDPQHTKKCFPAISHTSPVTLYDCSHKMGNDLWK 480  
Qy 549 YRKDKTYHPVSGSCMDSCSDHRIFMNTCNPSLTQOOWLEPHNISTYLEKFNEN 603  
Db 481 YRKDKTYHPVSGSCMDSCSDHRIFMNTCNPSLTQOOWLEPHNISTYLEKFNEN 535  
RESULT 10  
US-09-795-926-39  
; Sequence 39, Application US/09795926  
; Patent No. US20020098486A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sander, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Milganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kieke, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSEPERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/09/795,926  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 39  
; LENGTH: 366  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-795-926-39  
Query Match 57.3%; Score 1877; DB 9; Length 366;  
Best Local Similarity 100.0%; Pred. No. 9,5e-174;  
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 241 EAVVNMPLPDLDIRARRKTIIVCPMIDVIDHDDFRYETQAGDMRGAFFDWMYTKRIPIP 300  
Qy 301 PELQKADSPDPFSSPVWAGGLFAVDRKWFELGSDYDGLIHWGGEYISFKV 353  
Db 301 PELQKADSPDPFSSPVWAGGLFAVDRKWFELGSDYDGLIHWGGEYISFKV 353  
RESULT 11  
US-10-364-774-39  
; Sequence 39, Application US/10364774  
; Publication No. US20030144497A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sander, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Milganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kieke, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSEPERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/10/364,774  
; PRIOR FILING DATE: 2003-02-11  
; PRIOR APPLICATION NUMBER: US/09/795,926  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 39  
; LENGTH: 366  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-364-774-39  
Query Match 57.3%; Score 1877; DB 14; Length 366;  
Best Local Similarity 100.0%; Pred. No. 9,5e-174;  
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



RESULT 12  
US-10-292-896-3  
Sequence 3, Application US/10292896  
Publication No. US20030186850A1  
GENERAL INFORMATION:  
APPLICANT: HASSAN, Helle  
APPLICANT: REIS, Celso A.  
APPLICANT: BENNETT, Eric P.  
APPLICANT: CLAUSEN, Henrik  
TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GA  
TITLE OF INVENTION: TRANSEPERASS, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS  
FILE REFERENCE: 4305/1H154-US3  
CURRENT APPLICATION NUMBER: US/10/292,896  
PRIOR FILING DATE: 2003-03-17  
PRIOR APPLICATION NUMBER: US 60/425,204  
PRIOR FILING DATE: 2002-11-08  
PRIOR APPLICATION NUMBER: PCT/DK01/00328  
PRIOR FILING DATE: 2001-05-10  
PRIOR APPLICATION NUMBER: US 60/203,331  
PRIOR FILING DATE: 2000-05-11  
NUMBER OF SEQ ID NOS: 130  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 276  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-292-896-3

Query Match 42.5%; Score 1392; DB 14; Length 276;  
Best Local Similarity 97.3%; Pred. No. 1.2e-126;  
Matches 250; Conservative 3; Mismatches 2; Indels 2; Gaps 1;  
Qy 349 ISFKVMGGMEDIPGSRVGHIRKYVPYKVPAGVSLA--RNLKRVAVMMDEYAEIY 406  
Db 20 LAMVVMGGMEDIPGSRVGHIRKYVPYKVPAGVSLA--RNLKRVAVMMDEYAEIY 79  
Qy 407 ORRPEYRLHLSAGDVAVOKKRLSSLNCKSGFKMFMTKIADLPKPYPPVPPAANGELINV 466  
Db 80 ORRPEYRLHLSAGDVAVOKKRLSSLNCKSGFKMFMTKIADLPKPYPPVPPAANGELINV 139  
Qy 467 GTGCAIDTKGALSPRLBGCVRGREGAANNMVFPTFRREDIRPGDPQHTKKFCDA 526  
Db 140 GTGCAIDTKGALSPRLBGCVRGREGAANNMVFPTFRREDIRPGDPQHTKKFCDA 199  
Qy 527 ISHTSPVTLVYCHSMKGNQMLKRYRDKTLVHPVSSCMDCESDHRIEMTGNPSLTQQ 586  
Db 200 ISHTSPVTLVYCHSMKGNQMLKRYRDKTLVHPVSSCMDCESDHRIEMTGNPSLTQQ 259  
Qy 587 WLFEHINSTVLEKFRN 603  
Db 260 WLFEHINSTVLEKFRN 276

RESULT 13  
US-09-795-926-27  
Sequence 27, Application US/09795926  
Patent No. US20020098486A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedlich, Glenn  
APPLICANT: Abuin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Wiganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kieke, James Alvin  
APPLICANT: Potter, David George

TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
FILE REFERENCE: LEX-0144-USA  
CURRENT APPLICATION NUMBER: US/09/795,926  
CURRENT FILING DATE: 2001-02-28  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/186,558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 27  
LENGTH: 269  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-795-926-27

Query Match 41.8%; Score 1370; DB 9; Length 269;  
Best Local Similarity 100.0%; Pred. No. 1.6e-124;  
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 98 MTDARVDQAYRENGFNIVYSDKISLRSPLDIRHPNCNKRILETIPNTSIIIPFNEG 157  
Db 1 MTDARVDQAYRENGFNIVYSDKISLRSPLDIRHPNCNKRILETIPNTSIIIPFNEG 60  
Qy 158 WSSLRTVHSVYLSNPSPELVAEIVLVDFSDREHLKKPLDYMLPSPVAILRTKREG 217  
Db 61 WSSLRTVHSVYLSNPSPELVAEIVLVDFSDREHLKKPLDYMLPSPVAILRTKREG 120  
Qy 218 IRTMLGASVATGVDITFLDSHCANVNLPLDLRIARRKTIIVCMIDVIDHDDRYE 277  
Db 121 IRTMLGASVATGVDITFLDSHCANVNLPLDLRIARRKTIIVCMIDVIDHDDRYE 180  
Qy 278 TQADARAGAFDMYTKRIPIPELOKADPSDFESPVAAGLFAVDKRFMELGYDP 337  
Db 181 TQADARAGAFDMYTKRIPIPELOKADPSDFESPVAAGLFAVDKRFMELGYDP 240  
Qy 338 GLEIWMGEQYEISPKV 353  
Db 241 GLEIWMGEQYEISPKV 256

RESULT 14  
US-10-364-774-27  
Sequence 27, Application US/10364774  
Publication No. US20030144497A1  
GENERAL INFORMATION:  
APPLICANT: Donoho, Gregory  
APPLICANT: Hilbun, Erin  
APPLICANT: Turner, C. Alexander Jr.  
APPLICANT: Friedlich, Glenn  
APPLICANT: Abuin, Alejandro  
APPLICANT: Zambrowicz, Brian  
APPLICANT: Sands, Arthur T.  
APPLICANT: Walke, D. Wade  
APPLICANT: Wiganowski, Nathaniel L.  
APPLICANT: Hu, Yi  
APPLICANT: Kieke, James Alvin  
APPLICANT: Potter, David George  
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME  
FILE REFERENCE: LEX-0144-USA  
CURRENT APPLICATION NUMBER: US/10/364,774  
CURRENT FILING DATE: 2003-02-11  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US/09/795,926  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 60/185,920  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/186,558  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 60/191,849



;; PRIOR FILING DATE: 2000-03-24  
;; NUMBER OF SEQ ID NOS: 47  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 27  
;; LENGTH: 269  
;; TYPE: PRT  
;; ORGANISM: homo sapiens  
US-10-364-774-27

Query Match 41.8%; Score 1370; DB 14; Length 269;  
Best Local Similarity 100.0%; Pred. No. 1.6e-124;  
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDARVDAQAREGPNVYSDKISLNRLPDIRHPNCKRYLETLPNTSIIIPFNNEG 157  
DB 1 MTDARVDAQAREGPNVYSDKISLNRLPDIRHPNCKRYLETLPNTSIIIPFNNEG 60  
QY 158 WSSILRTVHSLVNSRPELVAEIVLVDDFSDREHLKKPLEDYMALPPSVRLRTKREGI 217  
DB 61 WSSILRTVHSLVNSRPELVAEIVLVDDFSDREHLKKPLEDYMALPPSVRLRTKREGI 120  
QY 218 IRTMLGASVATGVITFLDSHCANVMPLPDLRIARNKTIVCPMIDVIDHDFRYE 277  
DB 121 IRTMLGASVATGVITFLDSHCANVMPLPDLRIARNKTIVCPMIDVIDHDFRYE 180  
QY 278 TQAGDARGAPDMEMYYKRIPPELOKADSDPESVVMAGLPAYDRKKFELGYPD 337  
DB 181 TQAGDARGAPDMEMYYKRIPPELOKADSDPESVVMAGLPAYDRKKFELGYPD 240  
QY 338 GLBIWGEQVEISFKV 353  
DB 241 GLBIWGEQVEISFKV 256

RESULT 15  
US-09-795-926-35  
; Sequence 35, Application US/09795926  
; Patent No. US20020098486A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Miljanowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kieke, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/09/795,926  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 35  
; LENGTH: 321  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-795-926-35

Query Match 40.0%; Score 1311; DB 9; Length 321;  
Best Local Similarity 99.2%; Pred. No. 1.2e-118;  
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRRKREKLLQAVLVLAALVTLPLNVGIMALYREROPDGTGGSGAAVAPAGGSHSRK 60  
DB 1 MRRKREKLLQAVLVLAALVTLPLNVGIMALYREROPDGTGGSGAAVAPAGGSHSRK 60  
QY 61 KTFPLDGGQKLKQWHDKEAIRDAQRVNGEQGRPYPMTDAERYDQAYRENGFNIVYSDK 120  
DB 61 KTFPLDGGQKLKQWHDKEAIRDAQRVNGEQGRPYPMTDAERYDQAYRENGFNIVYSDK 120  
QY 121 ISLNRSLPDIRHPNCKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVNSRPELVAEI 180  
DB 121 ISLNRSLPDIRHPNCKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVNSRPELVAEI 180  
QY 181 VLVDFFSDREHLKKPLEDYMALPPSVRLRTKREGIIRRMIGASVATGVITFLDSHC 240  
DB 181 VLVDFFSDREHLKKPLEDYMALPPSVRLRTKREGIIRRMIGASVATGVITFLDSHC 240  
QY 241 EANNVMLPPLDGR 253  
DB 241 EANNVMLPPLDGR 253

RESULT 16  
US-10-364-774-35  
; Sequence 35, Application US/10364774  
; Publication No. US20030144497A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Miljanowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kieke, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/10/364,774  
; PRIOR FILING DATE: 2003-02-11  
; PRIOR APPLICATION NUMBER: US/09/795,926  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 35  
; LENGTH: 321  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-10-364-774-35

Query Match 40.0%; Score 1311; DB 14; Length 321;  
Best Local Similarity 99.2%; Pred. No. 1.2e-118;  
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRRKREKLLQAVLVLAALVTLPLNVGIMALYREROPDGTGGSGAAVAPAGGSHSRK 60  
DB 1 MRRKREKLLQAVLVLAALVTLPLNVGIMALYREROPDGTGGSGAAVAPAGGSHSRK 60  
QY 61 KTFPLDGGQKLKQWHDKEAIRDAQRVNGEQGRPYPMTDAERYDQAYRENGFNIVYSDK 120  
DB 61 KTFPLDGGQKLKQWHDKEAIRDAQRVNGEQGRPYPMTDAERYDQAYRENGFNIVYSDK 120  
QY 121 ISLNRSLPDIRHPNCKRYLETLPNTSIIIPFNNEGSSLLRTVHSLVNSRPELVAEI 180

Db 121 ISLRSLPDIHHPNCNSKRYLETLPNTSIIIPFHNEGWSLLRTVHSLVNSRPPELVAEI 180  
Qy 181 VLVDPSREHLKKPLEEDYMLPESVRLTRTKKRGILRTMTLGAASVATGVTIFLDSHC 240  
Db 181 VLVDPSREHLKKPLEEDYMLPESVRLTRTKKRGILRTMTLGAASVATGVTIFLDSHC 240  
Qy 241 EANYNMLPPLLD 253  
Db 241 EANYNMLPPLLGK 253

RESULT 17  
US-09-925-301-1006  
; Sequence 1006, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05882  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1006  
; LENGTH: 561  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-301-1006

Query Match 34.3%; Score 1125; DB 9; Length 561;  
Best Local Similarity 44.0%; Pred. No. 3,9e-100; Indels 38; Gaps 16;  
Matches 232; Conservative 87; Mismatches 170;

Qy 88 GNGEGRP--YPMTDAERVDQAYRENGFNIVYSDKISLNRSLPDIRHPNCNSKRYLETLP 145  
Db 59 GPGEMGRVVIPEKDOCKMEKFNQNLMASEMIALNRSLPDRLEGCKTKVYPNLP 118  
Qy 146 NTSIIIPFHNEGWSLLRTVHSLVNSRPPELVAEIVLVDPSREHLKKPLEEDYM-ALFP 204  
Db 119 TTSVVIYFHNEMAWSTLRTVHSLVNSRPHMLEIVLVDASERDFLKRPLESYVKKLV 178  
Qy 205 SVRLTRTKKRGILRTMTLGAASVATGVTIFLDSHCANYNMLPPLLDRIARNRKTYVCP 264  
Db 179 PVHIRMQRSGILRLRLKGAASVKGQVITFLDAHCECTGVWLEPLRLARIKHDRRTVCP 238  
Qy 265 MIDVIDHDDFRYETQAG--DAMRGAFDMEYTKRIPRP---ELQKADSPDPESPVMAG 320  
Db 239 IIVDISDTEFY--MAGSDMTYGGFNWKLNRMYVPQREMDRKGDTLPVRTPYMAAG 296  
Qy 321 LFAVDRKFMELGYDGLLEIWGGEQYEISFKVMCGRMEDIPCSRVGHYIRKTYPVYK 380  
Db 297 LFSIDRDFQELIGTYDAGMDIMGGENLEISFRIMQCGGLEIYVCSHVGHVFRKATPYTF 356  
Qy 381 PAGVS--LAARLKVAAEVMDEYAEIYQRRPEYRHLASGDVAVQKLRSLNSLCKSFKW 438  
Db 357 FGGTGQIINKNNRRLAEVWMDDEFKTFYIISPGVTKVYDGDISRGLRHKLQCRPFSWY 416  
Qy 439 MTKIAMD--LPKFPYVPEPPAAAMGEIRNVGTGLCAOTTKHGALGSLRLLEGCVRGGBAA 496  
Db 417 LENIYPSQIRRH-----FSLGEIRNVETNOCLDMMARKENKVGIFNC-HGMG--- 465  
Qy 497 WNNQVFTFTWRDIPRGPQHTKKCFDAISHTSPVTLVYDCHSKNGQMLKYRKDK-TL 555  
Db 466 --GNQVFSYTKANKIIRTD-----LCIDVSKLNGPVTMLKCHLKGQNLMEYDPVLT 517  
Qy 556 YHPVSGCMD-CSESDHRI-FMNTCNPSLSLTOQMLFHTNYSVLEKF 600  
Db 518 QHVNNSOCLDKATBEDSQVPSIRDCN-GSRSQQLLR--NVTLPBIF 561

RESULT 18  
US-10-001-851-24  
; Sequence 24, Application US/10001851  
; Publication No. US20020115628A1  
; GENERAL INFORMATION:  
; APPLICANT: MEYERS, Rachel A.  
; APPLICANT: WILLIAMSON, Mark  
; TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1 Human Glycosyl Transferase  
; FILE REFERENCE: 10147-5601  
; CURRENT APPLICATION NUMBER: US/10/001,851  
; PRIOR FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: US 60/249,939  
; PRIOR FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 559  
; TYPE: PRT  
; ORGANISM: Sus sp.  
US-10-001-851-24

Query Match 34.1%; Score 1117; DB 13; Length 559;  
Best Local Similarity 43.6%; Pred. No. 2.4e-99;  
Matches 230; Conservative 90; Mismatches 169; Indels 38; Gaps 16;

Qy 88 GNGEGRP--YPMTDAERVDQAYRENGFNIVYSDKISLNRSLPDIRHPNCNSKRYLETLP 145  
Db 57 GPGEMGRVVIPEKDOCKMEKFNQNLMASEMIALNRSLPDRLEGCKTKVYPNLP 116  
Qy 146 NTSIIIPFHNEGWSLLRTVHSLVNSRPPELVAEIVLVDPSREHLKKPLEEDYM-ALFP 204  
Db 117 TTSVVIYFHNEMAWSTLRTVHSLVNSRPHMLEIVLVDASERDFLKRPLESYVKKLV 176  
Qy 205 SVRLTRTKKRGILRTMTLGAASVATGVTIFLDSHCANYNMLPPLLDRIARNRKTYVCP 264  
Db 177 PVHIRMQRSGILRLRLKGAASVKGQVITFLDAHCECTGVWLEPLRLARIKHDRRTVCP 236  
Qy 265 MIDVIDHDDFRYETQAG--DAMRGAFDMEYTKRIPRP---ELQKADSPDPESPVMAG 320  
Db 237 IIVDISDTEFY--MAGSDMTYGGFNWKLNRMYVPQREMDRKGDTLPVRTPYMAAG 294  
Qy 321 LFAVDRKFMELGYDGLLEIWGGEQYEISFKVMCGRMEDIPCSRVGHYIRKTYPVYK 380  
Db 295 LFSIDRDFQELIGTYDAGMDIMGGENLEISFRIMQCGGLEIYVCSHVGHVFRKATPYTF 354  
Qy 381 PAGVS--LAARLKVAAEVMDEYAEIYQRRPEYRHLASGDVAVQKLRSLNSLCKSFKW 438  
Db 355 FGGTGQIINKNNRRLAEVWMDDEFKTFYIISPGVTKVYDGDISRGLRHKLQCRPFSWY 414  
Qy 439 MTKIAMD--LPKFPYVPEPPAAAMGEIRNVGTGLCAOTTKHGALGSLRLLEGCVRGGBAA 496  
Db 415 LENIYPSQIRRH-----FSLGEIRNVETNOCLDMMARKENKVGIFNC-HGMG--- 463  
Qy 497 WNNQVFTFTWRDIPRGPQHTKKCFDAISHTSPVTLVYDCHSKNGQMLKYRKDK-TL 555  
Db 464 --GNQVFSYTKANKIIRTD-----LCIDVSKLNGPVTMLKCHLKGQNLMEYDPVLT 515  
Qy 556 YHPVSGCMD-CSESDHRI-FMNTCNPSLSLTOQMLFHTNYSVLEKF 600  
Db 516 QHVNNSOCLDKATBEDSQVPSIRDCS-GSRSQQLLR--NVTLPBIF 559

RESULT 19  
US-10-001-851-21  
; Sequence 21, Application US/10001851  
; Publication No. US20020115628A1  
; GENERAL INFORMATION:  
; APPLICANT: MEYERS, Rachel A.  
; APPLICANT: WILLIAMSON, Mark  
; TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1 Human Glycosyl Transferase



US-10-205-219-76

Query Match 34.0%; Score 116; DB 14; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 2.9e-99;  
 Matches 230; Conservative 88; Mismatches 171; Indels 38; Gaps 16;

QY 88 GNGEGRP--YPMTDARVDOAYRENGFNIVSDKISLNSRLPDIRHNCNSKRYLETLP 145  
 DB 57 GPGEMGRPVVLPKEDQEKMEFKINQFNLMASEMIALNSRLPVRLEGCKTKYVPPNLP 116  
 QY 146 NTSIIIPFHNMGSSLIARTVSHVNSRPPELVAEIVLVDDPSDRHKKPLEDYM-ALFP 204  
 DB 117 TTSVVIHFHNHAWSTLIRTVSHVNSRPRHMEIEIVLVDDASERDFLKRLPLESVYKLV 176  
 QY 205 SVRLIRTKREGILRTMLGASVATGVITFLDSHCANVMWLPPLDIRIARNKTIYCP 264  
 DB 177 PVHVIIRMEQSGILIRALKGAIVSGVITFLDAHCECTVGMLEPLARIKHDRITVCP 236  
 QY 265 MIDVIDHDDFRYETQAG-DAMRGAFDWEWYKRIPIPP---ELQKADSPDFESPVMAG 320  
 DB 237 IIDVISDTFEY--MAGSDMTYGGFNWKLNFWMYFPVQREMDRRKGRDRTLPEVPTVMAG 294  
 QY 321 LEAIDRKFMELGIDPGLSIVGQVIEISFKWMCGRMEDIPCSRVGHITRYKYVYKV 380  
 DB 295 LFSIDRDFQBIIGTYDAGMDIWGGENLEISFRIMQCGGTLEIVTCSHVGFRRATPYTF 354  
 QY 381 PAGVS--LAARLKVAEVMDEVAEYIYQRRPEYRHLISAGDVAVOKKLRSLSNCKSKWF 438  
 DB 355 PGGTGQIINKNNRRLAEVWMDFNFFIISPGYTKVDYGDISRGVLRHLQCKPFSWY 414  
 QY 439 MTKIAMD--LPKFPVPEPPAAWGEIRNVGTGLCADTKGALGSPLRLBGCVRGGEAA 496  
 DB 415 LENIYPSQIPRHY-----FSLGEIRNVETNOCLDNMARKENKVGIFNC-HGMG--- 463  
 QY 497 WNNNOVFETFRREDIRPGDPQHTKKCFPDALSHTSPTLYDCHSKMGNQMLKRYKDK-TL 555  
 DB 464 --GNQVSYTANKSIRTD-----LCLDVSKLNGPVTMLKCHLKNQMLBEYDPVXLT 515  
 QY 556 YHPVSGCMD-CSESDEHRI-FMNTCNPSLSLTOQWLFHTNSTVLEKF 600  
 DB 516 QHVNNSOCLDKATEBDSQVPSIRDCT-GSRQOQMLR--NVTLPRIIF 559

RESULT 22  
 US-10-001-851-22  
 ; Sequence 22, Application US/10001851  
 ; Publication No. US20020115628A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MEYERS, Rachel A.  
 ; APPLICANT: WILLIAMSON, Mark  
 ; TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1el Human Glycosyl Transferase  
 ; FILE REFERENCE: 10147-5601  
 ; CURRENT APPLICATION NUMBER: US/0/001,851  
 ; CURRENT FILING DATE: 2001-11-20  
 ; PRIOR APPLICATION NUMBER: US 60/249,939  
 ; NUMBER OF SEQ ID NOS: 29  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 22  
 ; LENGTH: 559  
 ; TYPE: PRT  
 ; ORGANISM: Mus sp.  
 US-10-001-851-22

Query Match 34.0%; Score 115; DB 13; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 3.7e-99;  
 Matches 230; Conservative 87; Mismatches 172; Indels 38; Gaps 16;

QY 88 GNGEGRP--YPMTDARVDOAYRENGFNIVSDKISLNSRLPDIRHNCNSKRYLETLP 145  
 DB 57 GPGEMGRPVVLPKEDQEKMEFKINQFNLMASEMIALNSRLPVRLEGCKTKYVPPNLP 116

QY 146 NTSIIIPFHNMGSSLIARTVSHVNSRPPELVAEIVLVDDPSDRHKKPLEDYM-ALFP 204  
 DB 117 TTSVVIHFHNHAWSTLIRTVSHVNSRPRHMEIEIVLVDDASERDFLKRLPLESVYKLV 176  
 QY 205 SVRLIRTKREGILRTMLGASVATGVITFLDSHCANVMWLPPLDIRIARNKTIYCP 264  
 DB 177 PVHVIIRMEQSGILIRALKGAIVSGVITFLDAHCECTVGMLEPLARIKHDRITVCP 236  
 QY 265 MIDVIDHDDFRYETQAG-DAMRGAFDWEWYKRIPIPP---ELQKADSPDFESPVMAG 320  
 DB 237 IIDVISDTFEY--MAGSDMTYGGFNWKLNFWMYFPVQREMDRRKGRDRTLPEVPTVMAG 294  
 QY 321 LEAIDRKFMELGIDPGLSIVGQVIEISFKWMCGRMEDIPCSRVGHITRYKYVYKV 380  
 DB 295 LFSIDRDFQBIIGTYDAGMDIWGGENLEISFRIMQCGGTLEIVTCSHVGFRRATPYTF 354  
 QY 381 PAGVS--LAARLKVAEVMDEVAEYIYQRRPEYRHLISAGDVAVOKKLRSLSNCKSKWF 438  
 DB 355 PGGTGQIINKNNRRLAEVWMDFNFFIISPGYTKVDYGDISRGVLRHLQCKPFSWY 414  
 QY 415 LENIYPSQIPRHY-----FSLGEIRNVETNOCLDNMARKENKVGIFNC-HGMG--- 463  
 QY 497 WNNNOVFETFRREDIRPGDPQHTKKCFPDALSHTSPTLYDCHSKMGNQMLKRYKDK-TL 555  
 DB 464 --GNQVSYTANKSIRTD-----LCLDVSKLNGPVTMLKCHLKNQMLBEYDPVXLT 515  
 QY 556 YHPVSGCMD-CSESDEHRI-FMNTCNPSLSLTOQWLFHTNSTVLEKF 600  
 DB 516 QHVNNSOCLDKATEBDSQVPSIRDCT-GSRQOQMLR--NVTLPRIIF 559

RESULT 23  
 US-10-001-851-23  
 ; Sequence 23, Application US/10001851  
 ; Publication No. US20020115628A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MEYERS, Rachel A.  
 ; APPLICANT: WILLIAMSON, Mark  
 ; TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1el Human Glycosyl Transferase  
 ; FILE REFERENCE: 10147-5601  
 ; CURRENT APPLICATION NUMBER: US/10/001,851  
 ; CURRENT FILING DATE: 2001-11-20  
 ; PRIOR APPLICATION NUMBER: US 60/249,939  
 ; NUMBER OF SEQ ID NOS: 29  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 23  
 ; LENGTH: 559  
 ; TYPE: PRT  
 ; ORGANISM: Bos sp.  
 US-10-001-851-23

Query Match 34.0%; Score 115; DB 13; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 3.7e-99;  
 Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGEGRP--YPMTDARVDOAYRENGFNIVSDKISLNSRLPDIRHNCNSKRYLETLP 145  
 DB 57 GPGEMGRPVVLPKEDQEKMEFKINQFNLMASEMIALNSRLPVRLEGCKTKYVPPNLP 116  
 QY 146 NTSIIIPFHNMGSSLIARTVSHVNSRPPELVAEIVLVDDPSDRHKKPLEDYM-ALFP 204  
 DB 117 TTSVVIHFHNHAWSTLIRTVSHVNSRPRHMEIEIVLVDDASERDFLKRLPLESVYKLV 176  
 QY 205 SVRLIRTKREGILRTMLGASVATGVITFLDSHCANVMWLPPLDIRIARNKTIYCP 264  
 DB 177 PVHVIIRMEQSGILIRALKGAIVSGVITFLDAHCECTVGMLEPLARIKHDRITVCP 236  
 QY 265 MIDVIDHDDFRYETQAG-DAMRGAFDWEWYKRIPIPP---ELQKADSPDFESPVMAG 320

Db 237 IIDVDSDTREY--MAGSDMTYGGFNWKLNRWYVPVOREMDRRKGRDRTLPVPTPMAGG 294  
Qy 321 LFAVDRKMFELGSDYDGLIEWGEOYEISFKVMCGRMEDIDCSRGHLYRKYVYKV 380  
Db 295 LFSIDRNYFGEIGYDAGMDIWGEMLEISFRIMQCGGTLEIVTCSHVGHVFRATPYTF 354  
Qy 381 PAGVS--LAARLKRVAEVMDEVAEYIYORRPEYRHL SAGDVAVOKKLBSLACKSKPMF 438  
Db 355 PGGGGQIINKNNRRLAEVWMDERFKNFYIISPGVTKVDYDISRLRLHKLQCRPFSWY 414  
Qy 439 MTKIAMD--LPKFPYVPPAAAMGEBIRNVTGLCADTKHAGLSPRLBEGCVRGEEAA 496  
Db 415 LENIYPSQIPRHY-----FSLGEBIRNVTGCLDNMARENKNGKGFVNC-HGMG--- 463  
Qy 497 WNNMQUFTFTWRREDIRPGDPQHTKKFCFDAISHTSPVTLVDCHSMKNQMLKRYKDK-TL 555  
Db 464 --GNQVFSYANKKIRITDD-----LCIDVSKLNGPVTMLKCHLKGQMLWEYDPVXLTL 515  
Qy 556 YHPVSGSCMD-CSESDHRI--FNATCNPSLSLQOQLFHTNSTVLEKF 600  
Db 516 OHVNSNOCLDKATEBDSQVPSIRDCS-GSRSQOQLLR--NVTLPRIIF 559

## RESULT 24

US-10-001-851-20  
; Sequence 20, Application US/10001851  
; Publication No. US20020115628A1  
; GENERAL INFORMATION:  
; APPLICANT: MEYERS, Rachel A.  
; APPLICANT: WILLIAMSON, Mark  
; TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1el Human G[lycosyl] Transferase  
; TITLE OF INVENTION: Uses thereof  
; FILE REFERENCE: 10147-5601  
; CURRENT APPLICATION NUMBER: US/10/001, 851  
; CURRENT FILING DATE: 2001-11-20  
; PRIOR APPLICATION NUMBER: US 60/249, 939  
; PRIOR FILING DATE: 2000-11-20  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 559  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-001-851-20

Query Match 34.0%; Score 113; DB 13; Length 559;  
Best Local Similarity 43.6%; Pred. No. 5.8e-99;  
Matches 230; Conservative 87; Mismatches 172; Indels 38; Gaps 16;

Qy 88 GNGEOGRP--YPMTDARVDOAVRENGFNIVSDKISLNRSLPDIRHPNCKSKRYLETLP 145  
Db 57 GPGEMGKAVLIIPKDDQKMEKELKINQFNLMASSEMIALNLSLPVRLGCKTKYYPDLPL 116  
Qy 146 NTSIIIPFHNHNGSSLLRTYHVSYNRSPPRLVAEIVLVDDPSDRHLKKPLEDYMA-LFP 204  
Db 117 TTSVVAIFHNHNAVSTLRTYHVSYNRSPPRLVAEIVLVDDPSDRHLKKPLEDYMA-LFP 176  
Qy 205 SVRLTKRKKEGLIRTRMLGASVATGVTITFLDSHCANVNWLPPLDIRIARNRKTYVCP 264  
Db 177 PAVHIREGSGSLRRLKGAASVKGQVITFLDAHCECTVGMLEPLARIKHDRITVCP 236  
Qy 265 MIDVIHDDFRYETQAG-DAMRGAFDWMYKRIPIPP--ELQKADPSDPFSPVAVAG 320  
Db 237 IIDVDSDTREY--MAGSDMTYGGFNWKLNRWYVPVOREMDRRKGRDRTLPVPTPMAGG 294  
Qy 321 LFAVDRKMFELGSDYDGLIEWGEOYEISFKVMCGRMEDIDCSRGHLYRKYVYKV 380  
Db 295 LFSIDRNYFGEIGYDAGMDIWGEMLEISFRIMQCGGTLEIVTCSHVGHVFRATPYTF 354  
Qy 381 PAGVS--LAARLKRVAEVMDEVAEYIYORRPEYRHL SAGDVAVOKKLBSLACKSKPMF 438  
Db 355 PGGGGQIINKNNRRLAEVWMDERFKNFYIISPGVTKVDYDISRLRLHKLQCRPFSWY 414

Qy 439 MTKIAMD--LPKFPYVPPAAAMGEBIRNVTGLCADTKHAGLSPRLBEGCVRGEEAA 496  
Db 415 LENIYPSQIPRHY-----FSLGEBIRNVTGCLDNMARENKNGKGFVNC-HGMG--- 463  
Qy 497 WNNMQUFTFTWRREDIRPGDPQHTKKFCFDAISHTSPVTLVDCHSMKNQMLKRYKDK-TL 555  
Db 464 --GNQVFSYANKKIRITDD-----LCIDVSKLNGPVTMLKCHLKGQMLWEYDPVXLTL 515  
Qy 556 YHPVSGSCMD-CSESDHRI--FNATCNPSLSLQOQLFHTNSTVLEKF 600  
Db 516 OHVNSNOCLDKATEBDSQVPSIRDCN-GSRSQOQLLR--NVTLPRIIF 559

## RESULT 25

US-10-292-896-62  
; Sequence 62, Application US/10292896  
; Publication No. US20030186850A1  
; GENERAL INFORMATION:  
; APPLICANT: HASSAN, Helie  
; APPLICANT: REIS, Celso A.  
; APPLICANT: BENNETT, Eric P.  
; APPLICANT: CLAUSEN, Henrik  
; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE GAV  
; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENTS  
; TITLE OF INVENTION: SUCH AGENTS FOR PREPARING MEDICAMENTS  
; FILE REFERENCE: 4305/1H154-US3  
; CURRENT APPLICATION NUMBER: US/10/292, 896  
; CURRENT FILING DATE: 2003-03-17  
; PRIOR APPLICATION NUMBER: US 60/425, 204  
; PRIOR FILING DATE: 2002-11-08  
; PRIOR APPLICATION NUMBER: PCT/DK01/00328  
; PRIOR FILING DATE: 2001-05-10  
; PRIOR APPLICATION NUMBER: US 60/203, 331  
; PRIOR FILING DATE: 2000-05-11  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 62  
; LENGTH: 556  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-292-896-62

Query Match 33.7%; Score 1106; DB 14; Length 556;  
Best Local Similarity 43.8%; Pred. No. 2.8e-98;  
Matches 227; Conservative 90; Mismatches 165; Indels 36; Gaps 15;

Qy 88 GNGEOGRP--YPMTDARVDOAVRENGFNIVSDKISLNRSLPDIRHPNCKSKRYLETLP 145  
Db 56 GPGEMGKAVLIIPKDDQKMEKELKINQFNLMASSEMIALNLSLPVRLGCKTKYYPDLPL 115  
Qy 146 NTSIIIPFHNHNGSSLLRTYHVSYNRSPPRLVAEIVLVDDPSDRHLKKPLEDYMA-LFP 204  
Db 116 NTSVVAIFHNHNAVSTLRTYHVSYNRSPPRLVAEIVLVDDPSDRHLKKPLEDYMA-LFP 175  
Qy 205 SVRLTKRKKEGLIRTRMLGASVATGVTITFLDSHCANVNWLPPLDIRIARNRKTYVCP 264  
Db 176 PAVHIREGSGSLRRLKGAASVKGQVITFLDAHCECTVGMLEPLARIKHDRITVCP 235  
Qy 265 MIDVIHDDFRYETQAG-DAMRGAFDWMYKRIPIPP--ELQKADPSDPFSPVAVAG 320  
Db 237 IIDVDSDTREY--MAGSDMTYGGFNWKLNRWYVPVOREMDRRKGRDRTLPVPTPMAGG 293  
Qy 321 LFAVDRKMFELGSDYDGLIEWGEOYEISFKVMCGRMEDIDCSRGHLYRKYVYKV 380  
Db 294 LFSIDRNYFGEIGYDAGMDIWGEMLEISFRIMQCGGTLEIVTCSHVGHVFRATPYTF 353  
Qy 381 PAGVS--LAARLKRVAEVMDEVAEYIYORRPEYRHL SAGDVAVOKKLBSLACKSKPMF 438  
Db 354 PGGGGQIINKNNRRLAEVWMDERFKNFYIISPGVTKVDYDISRLRLHKLQCRPFSWY 413  
Qy 439 MTKIAMDLPKFPYVPPAAAM--GEIRNVTGLCADTKHAGLSPRLBEGCVRGEEAA 496  
Db 414 LENI-----YPSQIPRHYISLGEIRNVTGCLDNMARENKNGKGFVNC-HGMG--- 462

OY 497 WNNMVFETREDIRPGDPQHTKKCFDAISHSPVTLYDCHSKGNQLWKYRKD-TL 555  
 Db 463 --GNQVPSYTDKERTD-----LCUDVSRINGPVIMLKCHMRGNQLMEYDAERLTL 514  
 OY 556 YHPVSGSCMD-CSES DHRI-FMNTCNPSSLTQQLFEH 591  
 Db 515 RHVNSNOCLDEPSEEDKVPMTODCS-GSRSQOWILLRN 551

Search completed: November 22, 2004, 13:42:20  
 Job time : 58 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: November 22, 2004, 13:38:52 ; Search time 21 Seconds  
(without alignments)  
2762.796 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278  
Sequence: 1 MRKKEKRLQVALVLAALV.....TQGMLEHTNVTLEKRN 603

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :

1: PIR 79: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1392.5	42.5	684	2 T26930	hypothetical prote
2	1392	42.5	276	2 T12552	hypothetical prote
3	1125	34.3	559	2 JC4223	polypeptide N-acet
4	1115	34.0	559	2 A45987	polypeptide N-acet
5	1113	34.0	601	2 T42251	polypeptide N-acet
6	1083	33.0	624	2 T42247	polypeptide N-acet
7	1079	32.9	626	2 T42246	polypeptide N-acet
8	1078.5	32.9	623	2 T42245	probable polypepti
9	988.5	30.2	563	2 A88515	polypeptide N-acet
10	988.5	30.2	612	2 T42243	probable polypepti
11	968	29.5	617	2 T42249	polypeptide N-acet
12	949	28.9	571	2 T37405	polypeptide N-acet
13	948.5	28.9	618	2 T42248	polypeptide N-acet
14	938	28.6	562	2 T42250	polypeptide N-acet
15	934	28.5	633	2 JC5247	polypeptide N-acet
16	906	27.6	579	2 T31549	polypeptide N-acet
17	883	26.9	589	2 T42244	probable polypepti
18	801	24.4	605	2 T27357	hypothetical prote
19	642	19.3	421	2 T42252	polypeptide N-acet
20	147	4.5	1044	2 H97186	glycosyltransferas
21	139	4.2	306	2 D87531	glycosyl transfera
22	138.5	4.2	308	2 H87306	glycosyl transfera
23	136	4.1	328	2 B84253	electrolycan bios
24	135.5	4.1	312	2 S74669	hypothetical prote
25	135.5	4.1	318	2 D87506	glycosyl transfera
26	130	4.0	322	2 AC2023	hypothetical prote
27	130	4.0	787	2 H98163	hypothetical prote
28	128	3.9	470	2 C70641	hypothetical prote
29	126	3.8	313	2 A12404	hypothetical prote

30	124	3.8	362	2 G75191	dolichol-phosphate
31	124	3.8	972	2 T09595	glucuronosyltransf
32	123	3.8	334	1 G71153	hypothetical prote
33	122	3.7	316	2 AB2868	UDP-hexose transfe
34	122	3.7	316	2 F97644	UDP-hexose transfe
35	120	3.7	260	2 E90984	probable glycosyl
36	120	3.7	260	2 H85829	glycosyl transfera
37	119.5	3.6	416	2 AE1499	conserved hypotnet
38	118.5	3.6	298	2 B75096	glycosyl transfera
39	118.5	3.6	416	2 AB1141	probable glucosami
40	117.5	3.6	245	2 D87307	glycosyl transfera
41	117	3.6	343	2 A12091	glycosyltransferas
42	116	3.5	251	2 A12106	hypothetical prote
43	115.5	3.5	269	2 A13123	glycosyltransferas
44	115.5	3.5	332	2 H69814	hypothetical prote
45	115.5	3.5	477	1 US0589	endo-1,4-beta-xyla
46	114	3.5	333	2 AH2026	hypothetical prote
47	113.5	3.5	301	2 F95205	glycosyl transfera
48	111.5	3.4	211	2 A75176	dolichyl-phosphate
49	111.5	3.4	293	2 T00099	glycosyltransferas
50	111.5	3.4	383	2 T34603	xylosyltransferas
51	111	3.4	2774	2 A43359	microtubule-associ
52	110	3.4	215	1 D71038	hypothetical prote
53	108	3.3	233	2 H97103	probable glycosylt
54	108	3.3	274	2 B81319	probable glycosylt
55	108	3.3	353	2 H71223	probable dolichol-
56	108	3.3	448	2 T01814	hypothetical prote
57	107.5	3.3	322	2 T44647	glycosyl transfera
58	107.5	3.3	371	2 H69290	dolichol-P-glucose
59	105.5	3.2	319	2 A70715	hypothetical prote
60	105	3.2	294	2 D64175	glycosyltransferas
61	104.5	3.2	309	2 B87550	glycosyl transfera
62	104.5	3.2	581	2 B69322	dolichol-P-glucose
63	104.5	3.2	615	2 C97723	aspartate-tRNA lig
64	104.5	3.2	745	1 A70458	phosphoribosylform
65	104	3.2	277	2 C97626	dolichol-phosphate
66	104	3.2	279	2 AE2849	hypothetical prote
67	103.5	3.2	344	2 AC0974	probable glycosylt
68	103.5	3.2	348	2 D95977	glucosyltransferas
69	103.5	3.2	348	2 S39958	exoO protein - Rhl
70	103.5	3.2	420	2 D69769	cellulose synthase
71	103.5	3.2	570	2 S62627	agglutinin I precu
72	102.5	3.1	564	1 P1CSAG	agglutinin precurs
73	102	3.1	291	2 F93015	glycosyl transfera
74	102	3.1	295	2 AE2367	hypothetical prote
75	102	3.1	317	2 H97888	glycosyl transfera
76	102	3.1	328	2 F95158	glycosyl transfera
77	102	3.1	459	2 G69075	nitrogenase molybd
78	101.5	3.1	331	2 G75999	sugar transferase
79	101.5	3.1	344	2 A70037	capsular polysacch
80	101.5	3.1	335	2 A48755	hyaluronan synthas
81	101.5	3.1	419	2 A53100	hyaluronate synth
82	101	3.1	301	2 A81920	hypothetical prote
83	101	3.1	338	2 F98024	raffinose-raffinof
84	101	3.1	390	2 S76171	hypothetical prote
85	101	3.1	637	2 A28609	reticacycline resis
86	101	3.1	1143	2 S46122	SNF2 protein homol
87	100.5	3.1	318	2 AC2152	hypothetical prote
88	100.5	3.1	412	2 AC2152	dolichyl-phosphate
89	100.5	3.1	724	2 T04340	beta-galactosidase
90	100.5	3.1	1012	2 B90389	conserved hypotnet
91	100	3.1	307	2 AH1366	fructose-1-phosph
92	100	3.1	313	2 A12202	hypothetical prote
93	99.5	3.0	303	2 F81318	probable galactosy
94	99.5	3.0	1029	2 E85043	probable pre-mRNA
95	99	3.0	318	2 AG2189	hypothetical prote
96	99	3.0	327	2 G89990	hypothetical prote
97	99	3.0	475	2 T35697	arabinotransosidas
98	99	3.0	664	2 T47481	receptor-like prot
99	98.5	3.0	739	2 T12983	hypothetical prote
100	98	3.0	257	2 E84107	teichuronic acid b

## ALIGNMENTS

## RESULT 1

T26930  
hypothetical protein Y45F10D.3 - *Caenorhabditis elegans*

C/Species: *Caenorhabditis elegans*

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C/Accession: T26930

R/Murray A. submitted to the EMBL Data Library, January 1998

A/Reference number: Z20288

A/Accession: T26930

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-684 <MTL>

A/Cross-references: UNIPROT:O45947, EMBL:AL021492, PIDN:CAA16378.1; GSPDB:GN00022; CESP:

A/Experimental source: clone Y45F10D

C/Genetics:

A/Map position: 4

A/Intons: 61/3; 112/1; 142/3; 178/3; 230/3; 289/2; 611/2

C/Superfamily: polypeptide N-acetylglucosaminyltransferase

Query Match 42.5%; Score 1392.5; DB 2; Length 684;  
Best Local Similarity 48.8%; Pred. No. 1.6e-104;  
Matches 279; Conservative 87; Mismatches 175; Indels 31; Gaps 14;

```

QY 49 PAAGGSHSRQKTPFLDGGOKLKDMDHKEAIRDAQGVGEGGCRPTPTMTDAERVD-- 106
DB 121 PPALGGEALDPEFKYRGH-EKIK-MEDEAAYEKEKEGEGGCKPVKLPEDKEVEEA 178
QY 107 --AYRENGFNIVYSDKISLNSRLPDIRPNCNKKYLETLPNTSIIIPFHNGSSLLRT 164
DB 179 LSLYKANGYAYVYIDMSLNSIKDIRKECKNMYSKALPTVSITPFPHEHNSTLIRS 238
QY 165 VHSVLSNSPPELVVAIYLVDPDSDEHLKPLBDM--ALFPYVRLITRKRGRLRTR 221
DB 239 VYSVINSPPLELKEIILVDDPSEKPAALRQLEPLDKNGKIDHLYKVLRTKKRGLJGR 298
QY 222 MLGASVATGVTITLDSHCENANWMLPPLDLRIANRKTICPMIDVIDDHDFFRYETQAG 281
DB 299 QLGADATGELILFLDAISEANYWMLPPLDLPIEDYRTVCCPFVDVIDCEITYEVRP-D 357
QY 282 DAMGAFDMEMYKRIPIPELOKXADPPDPSPFMAAGLFAVDKRMWELGGYDPGLGI 341
DB 358 EGANGSEPMANVYKRLPLTK-DRESPTKPRNSPVMAGGYFAISKMWELGGIDEGLDI 416
QY 342 WGGEOYEISFYVMWCGGEMEDIPCSRGVHYR-KYVPYK-VPAGVSLANLKRVAEVMMD 399
DB 417 WGGEOYELSFYVMWCGGEMEDIPCSRGVHYR-KYVPYK-VPAGVSLANLKRVAEVMMD 476
QY 400 EYAAIYIQRRPEYHLSAGDVAVOGKLASSLNCSEFKKFKTKIAMDLPKFPVPEPPAAA 459
DB 477 DYKETLLYKRRGVGNADGDKMKMGIREKLCQCKSFDFMKEIADFDDKYPVAEPKSA 536
QY 460 WGEIRNVGTGCAADTKHGALGSPRLBEGCVR-----GRGEAAMNNVQVTFWREDIRGD 515
DB 537 EGEIRNVGTNCTIDTFKEQGNORGLAKCTSDKDGGE-----QDLRLRMHDIR-- 588
QY 516 POKTKKFCFDAISHT--SPVTLYDCHSMKNQMLWYR-KDXTLYHPVSGSCMDSESDDR 572
DB 589 --KGRKICFDCSTSVDAKAVILFDCHSMKNQMLFYKVAQIQIYHPISGQCLTADENKGG 646
QY 573 -IPNATCNPSLITQOMLEFHNSTYLEKFNRR 603
DB 647 FLHMKKCDSSSLQKMAWQTVDNELLETRQAN 678

```

## RESULT 2

T12552

hypothetical protein DKFZp586H0623.1 - human (fragments)

C/Species: *Homo sapiens* (man)

C/Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 09-Jul-2004

C/Accession: T12552

R/Amstrong, W.; Winkler, U.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, June 1999

A/Reference number: Z17527

A/Accession: T12552

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-150;151-276 <ANS>

A/Cross-references: UNIPROT:Q9Y4M4; EMBL:AL096739

A/Experimental source: adult uterus; clone DKFZp586H0623

A/Note: the cDNA sequence contains a -1 frameshift near codon 150

C/Genetics:

A/Note: DKFZp586H0623.1

C/Superfamily: polypeptide N-acetylglucosaminyltransferase

Query Match 42.5%; Score 1392; DB 2; Length 276;  
Best Local Similarity 97.3%; Pred. No. 4.6e-105;  
Matches 250; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

```

QY 349 ISFKVMWCGGEMEDIPCSRGVHYRKYVPYKVPAGVSLA--RNLKRVAEVMMDYAEYIY 406
DB 20 LAMQVMWCGGEMEDIPCSRGVHYRKYVPYKVPAGVSLARVTKRVAEVMMDYAEYIY 79
QY 407 QRRPEYHLSAGDVAVOGKLASSLNCSEFKKFKTKIAMDLPKFPVPEPPAAAAGEIRNV 466
DB 80 QRRPEYHLSAGDVAVOGKLASSLNCSEFKKFKTKIAMDLPKFPVPEPPAAAAGEIRNV 139
QY 467 GTGCAATTKHGALGSPRLBEGCVRGGEAAMNNVQVTFWREDIRGDPOHTKKFCFDA 526
DB 140 GTGCAATTKHGALGSPRLBEGCVRGGEAAMNNVQVTFWREDIRGDPOHTKKFCFDA 199
QY 527 ISHTSPVTLVDCHSMKNQMLWYRKDXTLYHPVSGSCMDSESDDRIPNATCNPSLITQ 586
DB 200 ISHTSPVTLVDCHSMKNQMLWYRKDXTLYHPVSGSCMDSESDDRIPNATCNPSLITQ 259
QY 587 WLFPHNSTYLEKFNRR 603
DB 260 WLFPHNSTYLEKFNRR 276

```

## RESULT 3

JC4223  
polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) - human

Alternate names: GalNAc-transferase

C/Species: *Homo sapiens* (man)

C/Date: 14-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 09-Jul-2004

C/Accession: JC4223; I37404

R/Meurer, J.A.; Naylor, J.M.; Baker, C.A.; Thomsen, D.R.; Homa, F.L.; Elhammer, A.P.

J. Biochem. 118, 568-574, 1995

A/Title: cDNA cloning, expression, and chromosomal localization of a human UDP-GalNAc: p

A/Reference number: JC4223; MUID:96115928; PMID:8690729

A/Accession: JC4223

A/Molecule type: mRNA

A/Residues: 1-559 <MEU>

A/Cross-references: UNIPROT:Q10472

A/Experimental source: salivary gland

A/Note: The authors translated the codon AAT for residue 264 as Asp

R/White, T.; Bennett, E.P.; Paul, E.; Takio, K.; Stensen, T.; Bonding, N.; Clausen, H.

J. Biol. Chem. 270, 24156-24165, 1995

A/Title: Purification and cDNA cloning of a human UDP-GalNAc:polypeptide N-Acetylglucto-

A/Reference number: I37404; MUID:96025800; PMID:7592619

A/Accession: I37404

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-559 <RES>

A/Cross-references: EMBL:X85018; NID:9971458; PIDN:CAA59380.1; PID:9971459

A/Comment: This enzyme catalyzes the initial reaction in O-linked (mucin type) oligosacchi-

retory pathway.

C/Genetics:

A/Map position: 18

C/Superfamily: polypeptide N-acetylglucosaminyltransferase

C/Keywords: chromosomal protein; glycoprotein; glycosyltransferase; hexosyltransferase;



F/9-28/Domain: transmembrane #status predicted <TM>

F/29-559/Domain: endoplasmic reticulum lumenal #status predicted <LM>

F/95-141,541,552/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F/117,118,288/Binding site: carbohydrate (Thr) (covalent) #status predicted  
F/119/Binding site: carbohydrate (Ser) (covalent) #status predicted

Query Match 34.0%; Score 1125; DB 2; Length 559;

Best Local Similarity 44.0%; Pred. No. 5,66-83;

Matches 232; Conservative 87; Mismatches 170; Indels 38; Gaps 16;

```

QY 88 GNGEQGRP--YPMTDARVQAVRENGFNIVSDKISLNSLPDIRHPNCKSKRYLETLP 145
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 57 GPGEMGRPVVYIPKEDQKMKEMFKINQFNLMASEMILNLSLPVRLGECCTKYYPNLP 116
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 146 NTSIIIPFNHNGWSSLRTVSHVSNRSPPELVAVIYVDDPSDEHLKKPLEDYM-ALFP 204
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 117 TTSVVIYFHNHNAVSTLIRTVSHVSNRSPPELVAVIYVDDASERDFLRPLESYVKGLKV 176
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 205 SVRLRTKREGLIRTMGLASVATGVITFLSHCANVWMLPPLDRIARAKTIVCP 264
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 177 PVHVRMEQSGILIRALKAASVKGQVITFLDHACECTGWLPEPLARIGHDKRTVCP 236
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 265 MIDVIDHDDPRYETQAG--DAMRGAFDWEMYYKRIPIPP---ELQKADPSDFESPVAAG 320
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 237 IIDVIDSDFEY--MAGSDMTYGGFNKKLNFRWYPVQREMDRKGRTLPVRIPTMAGG 294
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 321 LFAVDRKFMWELGSDYDGLIHWGEGEYISFKVWCGRMEDIPCSHVGHYRKTYVPYKV 380
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 295 LFSIDRDYFQEIIGYDAGMIMWGENLEISFRIMQCGGLEIYVCSHVGHVFRKATYTP 354
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 381 PAGVS--LAENLRVAEVMWDEYAEYIYQRRPEYRHLISAGDVAVQKTLRSLNCKSPKWF 438
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 355 PGGTGQIINKNNRRLAEVWMDDEFKNFFYIISPVTAKVDYGISRLGLRHLDQCRPSWY 414
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 439 MTKIAMD--LPKFPVPYPPAAAMGEIRNYGTGLCADTKGALGSLPLEGCVARGEEAA 496
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 415 LENIYPSQIPIRH-----FSLGEIRNVEYTNQCLDNMARKENKXGIFNC-HGMG--- 463
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 497 WNNMQVFTFWREDIRPGDPQHTKKCFDAISHTSPVTLVDCSHSKMNQMLKRYKDK-TL 555
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 464 --GNQVFSYANKKEIRTD-----LCIDVSKLNGPVTMLKCHHLKKNQMLWEYDPAVLT 515
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 556 YHPVSGSCMD-CSESDHRI-FMNTCNPSLSLQQLFPHNTSTVLEKF 600
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 516 QHVNNSQCLDKATVEDSQVPSIRDCN-GSRSQQMLLR--NVTLPETIF 559
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 4
A45987
polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C/Accession: A45987; A48530
R/Hagen, F.K.; Hollander, T.; Lehman, D.J.; Thomsen, D.R.; Elhammer, A.P.
J. Biol. Chem. 268, 12609-12616, 1993
A/Title: Isolation and expression of a cDNA clone encoding a bovine UDP-GalNAc:polypepti
A/Reference number: A45987; MUID:93286099; PMID:7685345
A/Accession: A45987
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-559 <HOM>
A/Cross-references: UNIPROT:Q07537; GB:L07780; NID:G289411; PIDN:AAA30532.1; PID:G289412
R/Note: authors translated the codon GAT for residue 310 as Asn
R/Hagen, F.K.; VanMuyckhuysen, B.; Tabak, L.A.
J. Biol. Chem. 268, 18960-18965, 1993
A/Title: Purification, cloning, and expression of a bovine UDP-GalNAc: polypeptide N-ac
A/Reference number: A48530; MUID:93366815; PMID:8360184
A/Accession: A48530
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 41-559 <HAG>
A/Cross-references: GB:L17437; NID:G304258; PIDN:AAA68489.1; PID:G304259
C/Superfamily: polypeptide N-acetylglactosaminyltransferase

```

C/Keywords: colostrum; glycosyltransferase; hexosyltransferase

Query Match 34.0%; Score 1115; DB 2; Length 559;

Best Local Similarity 43.6%; Pred. No. 3,66-82;

Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

```

QY 88 GNGEQGRP--YPMTDARVQAVRENGFNIVSDKISLNSLPDIRHPNCKSKRYLETLP 145
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 57 GPGEMGRPVVYIPKEDQKMKEMFKINQFNLMASEMILNLSLPVRLGECCTKYYPNLP 116
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 146 NTSIIIPFNHNGWSSLRTVSHVSNRSPPELVAVIYVDDPSDEHLKKPLEDYM-ALFP 204
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 117 TTSVVIYFHNHNAVSTLIRTVSHVSNRSPPELVAVIYVDDASERDFLRPLESYVKGLKV 176
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 205 SVRLRTKREGLIRTMGLASVATGVITFLSHCANVWMLPPLDRIARAKTIVCP 264
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 177 PVHVRMEQSGILIRALKAASVKGQVITFLDHACECTGWLPEPLARIGHDKRTVCP 236
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 265 MIDVIDHDDPRYETQAG--DAMRGAFDWEMYYKRIPIPP---ELQKADPSDFESPVAAG 320
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 237 IIDVIDSDFEY--MAGSDMTYGGFNKKLNFRWYPVQREMDRKGRTLPVRIPTMAGG 294
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 321 LFAVDRKFMWELGSDYDGLIHWGEGEYISFKVWCGRMEDIPCSHVGHYRKTYVPYKV 380
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 295 LFSIDRDYFQEIIGYDAGMIMWGENLEISFRIMQCGGLEIYVCSHVGHVFRKATYTP 354
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 381 PAGVS--LAENLRVAEVMWDEYAEYIYQRRPEYRHLISAGDVAVQKTLRSLNCKSPKWF 438
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 355 PGGTGQIINKNNRRLAEVWMDDEFKNFFYIISPVTAKVDYGISRLGLRHLDQCRPSWY 414
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 439 MTKIAMD--LPKFPVPYPPAAAMGEIRNYGTGLCADTKGALGSLPLEGCVARGEEAA 496
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 415 LENIYPSQIPIRH-----FSLGEIRNVEYTNQCLDNMARKENKXGIFNC-HGMG--- 463
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 497 WNNMQVFTFWREDIRPGDPQHTKKCFDAISHTSPVTLVDCSHSKMNQMLKRYKDK-TL 555
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 464 --GNQVFSYANKKEIRTD-----LCIDVSKLNGPVTMLKCHHLKKNQMLWEYDPAVLT 515
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 556 YHPVSGSCMD-CSESDHRI-FMNTCNPSLSLQQLFPHNTSTVLEKF 600
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 516 QHVNNSQCLDKATVEDSQVPSIRDCN-GSRSQQMLLR--NVTLPETIF 559
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 5
T42251
polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) 7 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: T42251
R/Hagen, F.K.; Nehrkne, K.
J. Biol. Chem. 273, 8268-8277, 1998
A/Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept
A/Reference number: 222126; MUID:98192620; PMID:955933
A/Accession: T42251
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-601 <HAG>
A/Cross-references: UNIPROT:O61397; EMBL:AF031841; NID:G3047202; PIDN:AA13677.1; PID:G3
A/Genes: gly-7
C/Superfamily: polypeptide N-acetylglactosaminyltransferase
C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 34.0%; Score 1113; DB 2; Length 601;
Best Local Similarity 42.0%; Pred. No. 5,86-82;
Matches 226; Conservative 90; Mismatches 186; Indels 36; Gaps 13;

QY 68 GQIKDMHDKAIRRDQKRVNGEQGRPYPMTDARV---DOAVRENGFNIVSDKISLN 124
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 79 GSEIGNYEPKEPEIPSNQ---PESHGKPVVTDDEGMAAGRAAEKFGFTVYSDMISN 135
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 125 RSLPDIRHPNCKSKRYLETLPNTSIIPFNHNGWSSLRTVSHVSNRSPPELVAVIYVDD 184
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Db 136 RTIPDIRPEBCKHWDYDEKLTPTVSVVVVFHNEGWTPLLRTHSVLRSPPLEIEQVVMVD 195  
Qy 185 DFDREHLKKKPLEDYMALF-PSVRLTRTKKEGLITRMLGASVATGVTFLDSHCEN 243  
Db 196 DSDKPHLKEKLDKYTRFNCKVIVTRQEBGLINRSISAKISTGEVFLAHGBVN 255  
Qy 244 VNNLPPLDIRARNKTIVCMDIVIDHDDFRYETOAGDA---MRGAFDWMYKRIPIIP 300  
Db 256 TNNLPPLAIKRRKRWMTVVIDIGDINSWEYSVSGPNAHSGIFEMWLKYEQIT 315  
Qy 301 PE--LQCADSPDPSPVMAAGLPAVDRKFWELGCDPGLTNGSGOYEISFVVMCGG 358  
Db 316 ERETAHRKHSQPRSPTHAGGLPAIRNLMPKELGYDEGIQIWMGGQYELSFIMQCGG 375  
Qy 359 RMEDIPCSRVGHIYRKVYVY---KVPAGVSLARULKVAEVMDEAYIYQRPPEYRL 415  
Db 376 GIVFVPCSHVGHVTRSHMPSPFGKSPGVISINMRVVKTMWDYISKYITREPOATNV 435  
Qy 416 SAGDAVAVOKLRSLSLNCKSPKMTKIAMDLPKFPVPEPAAAMGEIRNVGTGLCADTK 475  
Db 436 NPGDISQLLRDLCKSPKMTMENVAVDLKSYPML-PENDVWGEARNPATCKLD-R 493  
Qy 476 HGALGSPRLREGCVRGGEAANNMNVFTFWREDIRPDQHTKKCFDAISHTSPVTL 535  
Db 494 MGGIPGEMGATGC-----HGYGNQLIRLNVQGMQGE-----MCLTR---NGIRI 537  
Qy 536 YDCHSMKG--NQLKRY-RKDKTLVHPVSGCMDSSEDRIFMNTCNPSLSLTOQMLPE 590  
Db 538 QANHCVGTGVGFWSYDRKTKQIHSQKQCIIVSESGEVTLDTCTEDNEROKFWK 595

## RESULT 6

T42247  
polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T42247  
J. Hagen, F.K.; Nehrke, K.  
R. Biol. Chem. 273, 8268-8277, 1998  
A/Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:polypep-  
A/Reference number: Z22126; MUID:98192620; PMID:9525933  
A/Accession: T42247  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-624 <HAG>  
A/Cross-references: UNIPROT:O61393; EMBL:AF031837; NID:G3047194; PIDD:AA03673.1; PID:G3  
C/Genetic8:  
A/Gene: gly-5  
C/Superfamily: polypeptide N-acetylgalactosaminyltransferase  
C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 33.0%; Score 1083; DB 2; Length 624;  
Best Local Similarity 45.0%; Pred. No. 1.7e-79;  
Matches 226; Conservative 76; Mismatches 166; Indels 34; Gaps 13;  
Qy 103 RVDAVYENGFNIVSDKISLNRSLPDIRHPNCSKRYLETLPTNTSIIIPHNMGWSSL 162  
Db 133 KYDGMNMFNOYASDMISVHRTLPTNIDAECKTEKNENLPTSVITICFHNEMAVTL 192  
Qy 163 RTVSHVNRSPPELVAEIVLVDDPSDREHLKKPLEDYMALF-PSVRLTRTKKEGLIRTR 221  
Db 193 RTVSHVNRSPPELVAEIVLVDDPSDREHLKKPLEDYMALF-PSVRLTRTKKEGLIRTR 221  
Qy 222 MLGASVATGVITFLDSHCENANWMLPRLDIRANKRTIYCPMTDIVDDHDFY-ETQA 280  
Db 253 LRGAAVAGLEVLTLDSCHECMGEMELDLRIKRDPTTVVCPVIDIDDTTFEYHNSKA 312  
Qy 281 GDARGAFAFDEMYKRIPIPELOK--ADSPDPSPVMAAGLPAVDRKFWELGCDP 338  
Db 313 YFTSVGGFDNGLQGNHMSIPERDRKRRTPIDPVRSTPMAGGLSIDKEYEKGLTDPG 372  
Qy 339 LEIWGSGOYEISFVVMCGGEMEDIPCSRVGHIYRKVYVYKVPAGVS-LARNLKVAVEW 397  
Db 509 EDIRPDQHTKCFDAISHTSPVTLYDCHSMKGNOLMKYRKDK-TLYHPVSGSCMDSC 567

Db 373 FDIWGENLELSEFKIMWCGGTLEIVPCSHVGHVFRKSPYKMTGTVNLKRNIRLAEVW 432  
Qy 398 MDEYAEIYQRPPEYRLSAGDAVAVOKLRSLSLNCKSPKMTKIAMDLPKFPVPEP 457  
Db 433 LDDKYTYIERINNOGL-DIGDISSRKCLREDLQCSFKMYLONI---YBELFVPGE--S 486  
Qy 458 AAMGEIRNVGTGLCADTKHGALGSPRLREG---CYRGGEAANNMNVFTFWREDIR 512  
Db 487 VAKGELRNAQTSQCLDS---AVGEVENKAITTPYPCHEQGNQY-----MLSKGEIR 537  
Qy 513 PGPQHTKKCFDAISHTSPVTLYDCHSMKGNOLMKYRKDK-TLYHPVSGSCMDSCSDH 571  
Db 538 RDES-----CVDYAG--SDVMVFPCHGMKGNDEMRYNHDTGRLOHRAVSQKCLAMTXDGA 589  
Qy 572 RIFMNTCNPSLSLTOQMLPEHTN 593  
Db 590 KLENVACQYDDPYOHMKFKEYN 611

## RESULT 7

T42246  
polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T42246  
J. Hagen, F.K.; Nehrke, K.  
R. Biol. Chem. 273, 8268-8277, 1998  
A/Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:polypep-  
A/Reference number: Z22126; MUID:98192620; PMID:9525933  
A/Accession: T42246  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-626 <HAG>  
A/Cross-references: UNIPROT:O61392; EMBL:AF031836; NID:G3047192; PIDD:AA03672.1; PID:G3  
C/Genetic8:  
A/Gene: gly-5  
C/Superfamily: polypeptide N-acetylgalactosaminyltransferase  
C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 32.9%; Score 1079; DB 2; Length 626;  
Best Local Similarity 44.9%; Pred. No. 3.5e-78;  
Matches 227; Conservative 75; Mismatches 164; Indels 40; Gaps 14;  
Qy 103 RVDAVYENGFNIVSDKISLNRSLPDIRHPNCSKRYLETLPTNTSIIIPHNMGWSSL 162  
Db 133 KYDGMNMFNOYASDMISVHRTLPTNIDAECKTEKNENLPTSVITICFHNEMAVTL 192  
Qy 163 RTVSHVNRSPPELVAEIVLVDDPSDREHLKKPLEDYMALF-PSVRLTRTKKEGLIRTR 221  
Db 193 RTVSHVNRSPPELVAEIVLVDDPSDREHLKKPLEDYMALF-PSVRLTRTKKEGLIRTR 221  
Qy 222 MLGASVATGVITFLDSHCENANWMLPRLDIRANKRTIYCPMTDIVDDHDFY-ETQA 280  
Db 253 LRGAAVAGLEVLTLDSCHECMGEMELDLRIKRDPTTVVCPVIDIDDTTFEYHNSKA 312  
Qy 281 GDARGAFAFDEMYKRIPIPELOK--ADSPDPSPVMAAGLPAVDRKFWELGCDP 338  
Db 313 YFTSVGGFDNGLQGNHMSIPERDRKRRTPIDPVRSTPMAGGLSIDKEYEKGLTDPG 372  
Qy 339 LEIWGSGOYEISFVVMCGGEMEDIPCSRVGHIYRKVYVYKVPAGVS-LARNLKVAVEW 397  
Db 373 FDIWGENLELSEFKIMWCGGTLEIVPCSHVGHVFRKSPYKMTGTVNLKRNIRLAEVW 432  
Qy 398 MDEYAEIYQRPPEYRLSAGDAVAVOKLRSLSLNCKSPKMTKIAMDLPKFPVPEP 457  
Db 433 LDDKYTYIERINNOGL-DIGDISSRKCLREDLQCSFKMYLONI---YBELFVPGE--S 486  
Qy 458 AAMGEIRNVGTGLCADTKHGALGSPRLREG---CYRGGEAANNMNVFTFWREDIR 512  
Db 487 VAKGELRNAQTSQCLDS---AVGEVENKAITTPYPCHEQGNQY-----MLSKGEIR 537  
Qy 513 PGPQHTKKCFDAISHTSPVTLYDCHSMKGNOLMKYRKDK-TLYHPVSGSCMDSCSDH 571  
Db 538 RDES-----CVDYAG--SDVMVFPCHGMKGNDEMRYNHDTGRLOHRAVSQKCLAMTXDGA 589  
Qy 572 RIFMNTCNPSLSLTOQMLPEHTN 593  
Db 590 KLENVACQYDDPYOHMKFKEYN 611

Db 532 LS-KDGEIRDES-CVDYAG--SDVWV/PCHGKMGNOEMRYNHDTRGLQHAVSQKLGMT 587

Qy 568 BSDHRIFMANTCNPSSLTQOMLFEHTN 593

Db 588 KDGAKLEMAVCQYDDPYQHWKFKSYN 613

# RESULT 8

T42245

probable polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) - Caenorhabditis el

C:Species: Caenorhabditis elegans

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: T42245

R:Hagen, R.K.; Nehrke, K.

J. Biol. Chem. 273, 8268-8277, 1998

A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Poly

A:Reference number: Z22126; MUID:98192620; PMID:955933

A:Accession: T42245

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-623 <HAG>

A:Cross-references: UNIPROT:O61391; EMBL:AF031835; NID:g3047190; PIDN:AAJ3671.1; PID:g3

A:Gene: g3ly-5

C:Superfamily: polypeptide N-acetylglactosaminyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 32.9%; Score 1078.5; DB 2; Length 623;

Best Local Similarity 45.3%; Pred. No. 3.8e-79;

Matches 226; Conservative 74; Mismatches 170; Indels 29; Gaps 13;

Qy 103 RVQDAIRENPNIVYSKISLNSLPDIRHNCNKSKYLETLPNTSIIIPHNAGSSLL 162

Db 133 KYDKGMLNNAFNOYASDMISVHRTLPTNIDAECKTEKYNELPTSYIIFHNEMASVLL 192

Qy 163 RTVHSVLRNPPELVAETVVDPSREHLKKPLEDMALF-PEVRLTRTKRGELRTR 221

Db 193 RTHSVLERTPDHLLEEVLDVDDSDMDHTRPLEEYMSQFGKVKILRMEKREGLRRAR 252

Qy 222 MLAGASVATGVTITFLDSHCANVNMPLDLRIARNKTIYCPMIDVIDHDDEFRY-ETOA 280

Db 253 LRGAATAVATGVLTYLDHSCOMEGMEPRLDRIKRDPTTVVCPVIDVIDNTPEYHNSKA 312

Qy 281 GDAMKAFDEWETYYKRIPIPELQK--ADSPDPESPVMAGGLFPAVDRKMFELGYDPG 338

Db 313 YFTSVGFGFDWGLQFMNHSIPERDKNTRPRDPVRSPTMAGLFSIDKEFEKLGTYDPG 372

Qy 339 LETWGEQVYISFVMMCGSRMEDIPCSRYGHIYRKVPYKVPAGVS-LARNLKRVAEW 397

Db 373 FDIWGGNLELSPKIMWGGTLEIVPCSHVGHVFRKSPYKMTGTVNLKNSIRLAEVW 432

Qy 398 MDEYAEIYQRRPEYRHLASADVAVQKLRSLNCKSPKWFMTKIADLPRFYPPVPPA 457

Db 433 LDDYKTYVYRINNQLG-DFGDISSRKGLREDLGCKFSKMYLDNI---YELFLPVG--S 486

Qy 458 AAMGEIRNVG--TGLCADYTHGALGSPRLLEGCVRGGEAANNMVOYFTTWREDIRPD 515

Db 487 VAKGEMNNAGKQRQCIDYKPSG-GKTVGYQCNQGNQYV-----MLSKGEIRDRS 539

Qy 516 PQHTKKCFPAISHTSPVTLTYDCHSMKGNOLMKYRKQ-TLYHPVSGSCMDCSGSHRIF 574

Db 540 S-----CVDYAG--SDVWV/PCHGKMGNOEMRYNHDTRGLQHAVSQKLGMTGAKLE 591

Qy 575 MNTCNPSLSTQOMLFEHTN 593

Db 592 MVACQYDDPYQHWKFKSYN 610

# RESULT 9

A88515

polypeptide N-acetylglactosaminyltransferase homolog - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 02-Nov-2001

C:Accession: A88515

R:Anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog

A:Reference number: A75000; MUID:99069613; PMID:9851916

A>Note: see webstites genome.wustl.edu/genec/ C. elegans/ and www.sanger.ac.uk/Projects/C. ele

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A:Accession: A88515

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-563 <STO>

A:Cross-references: GB:chr\_III; PIDN:AAA28224.1; PID:g289776; GSPDB:GN00021

C:Genetics:

A:Map position: 3

C:Superfamily: polypeptide N-acetylglactosaminyltransferase

Query Match 30.2%; Score 988.5; DB 2; Length 563;

Best Local Similarity 40.5%; Pred. No. 6.5e-72;

Matches 211; Conservative 87; Mismatches 182; Indels 41; Gaps 15;

Qy 88 GNGEGRPYPMTDAERVDAVRENGENIYVSDKISLNSLPDIRHNC-NSKRYLET--L 144

Db 61 GGGGTGTVPEDEKTKIKKFLKQFNVASEMISVHRTLPTNIDAECKTEKYNELPTSYIIFHNEMASVLL 192

Qy 145 PNTSIIIFHNEMASVLLRTLSHSVINSPRHLEIILVDKSDRDYLVRLDSYIMQFP 180

Db 121 PKTSIIIFHNEMASVLLRTLSHSVINSPRHLEIILVDKSDRDYLVRLDSYIMQFP 180

Qy 205 -SVRIITKREGILRTIRMLGASVATGVTITFLDSHCANVNMPLDLRIARNKTIYCP 263

Db 181 IPIHVLNRSGLIRARLGSSEMAKGLILFDHAYEVNDGMLRPLVSVADERKVVVA 240

Qy 264 PMIDVIDHDDEFRYETQAGDMKRAFDEWETYYKRIPIPE-ELQK--ADSPDPESPVMAGG 320

Db 241 PIIIVISDTPFEYVLT-ASETTWGGFNMHLNFRWAVAPKRELNRGSDRSMPIQTPTIAGG 299

Qy 321 LFAVDRKMFELGYPGILEINGEQEISFKYMMCGSRMEDIPCSRYGHIYRKVPYKVP 380

Db 300 LFAIDKQFFYDLSGYBGMWGGEMLEISFRVWCGSLIHLPCSVGHVFRKQTYTF 359

Qy 381 PAGVS--LAANLKRVAEWMDEYAEIYQRRPEYRHLASADVAVQKLRSLNCKSPKMF 438

Db 360 PGSTAKYIHNAARTAEVWDEYKAPFYKVPAAKRVADVSERKGLRETQCKSKMY 419

Qy 439 MTKIAMDLPKRYPPVEPPA--AAMGEIRNVGTGLCADYTHGALGSPRLLEGCVRGGEA 496

Db 420 LENI-----YEPALPLADFRSLGAILVNRFTKCVDTNGKKGQAPGIDACHGAGNOA 472

Qy 497 WNNMVOYFTTWREDIRPDQHTKKCFPAISHT---SEVTLTYDCHSMKGNOLMKYRKQ 552

Db 473 W-----SLTGKGEIRSD-----LCLSS-GHYTOIGSELKLERCSVKINVGHVFD 519

Qy 553 K---TLYHPVSGSCMDCSGSHRIFMNTCNPSLSTQOMLE 590

Db 520 DQAGTILHKTGKCV--TGADQVTLTDECGLGKRDQMKQLE 558

# RESULT 10

T42243

probable polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) - Caenorhabditis el

C:Species: Caenorhabditis elegans

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: T42243

R:Hagen, R.K.; Nehrke, K.

J. Biol. Chem. 273, 8268-8277, 1998

A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Poly

A:Reference number: Z22126; MUID:98192620; PMID:955933

A:Accession: T42243

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-612 <HAG>

A:Cross-references: UNIPROT:P34678; EMBL:AF031833; NID:g3047186; PIDN:AAJ3669.1; PID:g30

R:Wilson, R.

submitted to the EMBL Data Library, May 1993  
A:Description: Sequence of the C. elegans coemid ZK688.

A:Reference number: 544913

A:Accession: 544913

A:Molecule type: DNA

A:Residues: 50-612 <MIL>

A:Cross-references: EMBL:L16621; NID:G289775; PIDN:AAA28224.1; PID:G289776

C:Genetics:

A:Gene: g1y-3

A:Insertion: 101/1; 229/3; 520/3; 570/3

C:Superfamily: polypeptide N-acetylglucosaminyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 30.2%; Score 988.5; DB 2; Length 612;  
Best Local Similarity 40.5%; Pred. No. 7.4e-72;  
Matches 211; Conservative 87; Mismatches 182; Indels 41; Gaps 15;

Qy 88 GNGGQRPYPTMDERDQAVRENGFNIVYSDKISLRSLPDIDHPNC-NSKRIYER-L 144  
Db 110 GGGGTGVTPEDKKTIRKRELENOFNVASSEMISVNRTPDYSDACRTSGNNLKTAGM 169

Qy 145 PNTSIIIPFNHSGSSILRTVSHVSNRSPPELVAEIVLVDPFSREHLKPELDYMLFP 204  
Db 170 PKTSIIIVFNHNAETTLRTIASHVINSFRHLEIILVDDKSDRDYLVKRLDYIKMP 229

Qy 205 -SVRIETKRGELIRTRMLGASVATGVDITFLDSHCANVMVLPILDIRIARRKTIYC 263  
Db 230 IPIHLVLENRSGILIRARLTGSEMAKGIILFLDAHVEVTDGWLPLVSRVAEDRKIVVA 289

Qy 264 PMIDVIVHDDPRYTOAGDMRGAPEMRYKRIPIRP-ELQK--ADSPDFESPVMAG 320  
Db 290 PIIIDVIDDTPFEYVT-ASETTWGGFNMHINFRMYAVPGRGLNRRGSDRSMPIQPTIAGG 348

Qy 321 LFAVDRKMFELGSDYDGLIWMGEQYEISFKVMCGRMEDIPCSRGIHYRYKVPYKV 380  
Db 349 LFAIDKQPTIDIGSDYDEMQWGENLEISFRVMCGSLIHCSSRGVHFRKQTPYTF 408

Qy 381 PAGVS--LARNLKAEVAMDEVAEYIYORRPEYRHLISAGDAVAVQKLRSSLNCKSEKMF 438  
Db 409 PGGAFAKVIHNAATAEVAMDEYKAFYKQVPAARVAGVSRKRLRETLQCKSPKMY 468

Qy 439 MTKIAMLPKRYPEVPEPA--AANGELRNVTGICADTGHGALSPRLBECRGREGAA 496  
Db 469 LENT-----YPRAPLPADFRSLGAIYNRTEKCVDTNGKDDQAPCIQACHGAGGQA 521

Qy 497 MNNAQVFTWRBEDIRPDPOHTKKCFDAISHT-----SPVLYDCHSMKGNQLMKYRKD 552  
Db 522 W-----SLTGKGEIRSD-----LCLSS-GHYVQIGSELKLRCSVSKINXHVFPD 568

Qy 553 K--TLVHPVSGSCMDCSDHRIEMTNCPSILTOQLFE 590  
Db 569 DQAGTLHKTKGKCV--TGADQVRLTDECGLRGKQWQMLE 607

RESULT 11  
T42249  
polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) 6b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: T42249

R:Hagen, F.K.; Nehrtke, K.

J. Biol. Chem. 273, 8268-8277, 1998

A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:Polypept

A:Reference number: Z22126; MUID:98192620; PMID:9525933

A:Accession: T42249

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-617 <HAG>

A:Cross-references: UNIPROT:O61395; EMBL:AF031839; NID:G3047198; PIDN:ANCI3675.1; PID:G3

C:Genetics:

A:Gene: g1y-6

C:Superfamily: polypeptide N-acetylglucosaminyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 29.5%; Score 968; DB 2; Length 617;  
Best Local Similarity 39.3%; Pred. No. 3.4e-70;  
Matches 216; Conservative 96; Mismatches 162; Indels 76; Gaps 22;

Qy 73 DMHDKAIIRRDQAVRGNGEGRPYPTMDARV--DQAVRENGFNIVYSDKISLRSLPDIDR 131  
Db 97 DW-----GGGAGVSHLTPEDKKTIRKRELENOFNVASSEMISVNRTPDYSDACRTSGNNLKTAGM 143

Qy 132 HPNCSKRYLETLPNTSIIIPFNHSGSSILRTVSHVSNRSPPELVAEIVLVDPFSRDL 191  
Db 144 KPCSRMNTYDNLPTTSIIIVFNHNAETTLRTIASHVINSFRHLEIILVDDKSDRDYLVKRLDYIKMP 203

Qy 192 LKRP-LEDYMLPPS-VRIETKRGELIRTRMLGASVATGVDITFLDSHCANVMVLP 249  
Db 204 LRYPTLDTLTKLPPTDIIKIRSKRVGLIRARMTGAQEOGVDVITFLDSHCCECTKGLP 263

Qy 250 LDRIRANRRTIYOPMIDVIDHDDPRYTOAGDMRGAPEMRYKRIPIRP-ELQK--A 306  
Db 264 LTRIKLNRAVPCPVLDIINDTFOY-KGIEFRGCFNMNLOFRMYGMPYAKOHL 322

Qy 307 DPSPDFESPVMAGGLFAVDRKMFELGSDYDGLIWMGEQYEISFKVMCGRMEDIPCS 366  
Db 323 DPICPISPTMAGGLFENNINYPBELGEYDPCMDINGENLEMSFRIMCGGRELIPCS 382

Qy 367 RVGHIYKRYVPYKVP--AGVSLARNLKAEVAMDEVAEYIYORRPE-YRHLISAGDAV 422  
Db 383 HVGVFPRKSSPHDPFGKSGKVLNTNLRLVAEVMDDMKYFYKIALQARHMSISIVSE 442

Qy 423 QKLRSSLNCKSPFEMTKIAMLPKRYPEVPEPAAMGHIRVNTGLC-----ADT--- 474  
Db 443 RVELRKLNCKSPFMYIQANVPQD--HFLP--TFLDRFGKSN--SNYCTAFRPGDTPK 495

Qy 475 KHGALGSPRLBLEGCVRGGAAMNMQVFTWRBEDIRPDPOHTKKCFDAIS--HTS- 531  
Db 496 NHRLLGSP-----CTMG-----FDLMQMLYTGRRIRTE-----HLCLSVGLHTTS 540

Qy 532 --PVTLYDCHSMKGNQLMKYR-KDKTLYHPVSGSCMDCSDHRIP-----MNT 577  
Db 541 DMKIQLEKAGF--DTEYWDKPKIGRFQNRKTCGLASPD----IFPTKDFENPPIYOK 595

Qy 578 CNPSILTOQM 587  
Db 596 CRSSNDROW 605

RESULT 12  
I37405.  
polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) - human

C:Species: Homo sapiens (man)

C:Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004

C:Accession: I37405

R:White, T.; Bennett, E.P.; Paul, E.; Taklo, K.; Stensen, T.; Bonding, N.; Clausen, H.

J. Biol. Chem. 270, 24156-24165, 1995

A:Title: Purification and cDNA cloning of a human UDP-GalNAc:polypeptide N-Acetylglucto

A:Reference number: I37404; MUID:96025800; PMID:7592619

A:Accession: I37405

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-571 <RES>

A:Cross-references: UNIPROT:Q10471; EMBL:X85019; NID:G971460; PIDN:CAA59381.1; PID:G9714

C:Genetics:

A:Gene: GDB:GALNT2; GALNAC

A:Cross-references: GDB:696223

A:Map position: 16q24-16q24

C:Superfamily: polypeptide N-acetylglucosaminyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 29.0%; Score 949; DB 2; Length 571;  
Best Local Similarity 33.5%; Pred. No. 1.1e-68;  
Matches 221; Conservative 103; Mismatches 170; Indels 166; Gaps 21;

Qy 1 MRRERKLLQAVLVALVLAIVLPVNGIMLYRERQDPTGGGAAVAPAAQGGSHSRQK 60

```

Db      1 MRRSRMLTLCFAFLWLVGIA-----YNY-----SCGSSALAGAAGCGA----- 39
Qy      61 KTFPLDGOQKLDKMDHXAIRDAQVNGEQ-----GR-PYPMTDARV----- 104
Db      40 -----GRK-EDMNEIDPIKKDKDLSHNGEKAQSMETLPPGKRVMPDFNOEAUYVGTM 91
Qy      105 -----DQYRNGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEGMS 160
Db      92 VRSQGDPRYAKRFQVESDKLRMDRAIPDTHDQCRQKQWVDLPATSVVTFPNEARSA 151
Qy      161 LRTVHSLNRSPEPLVAEIVLVDFSDREHLKKPLD--YMAFPSPVRLITKKRGLI 218
Db      152 LRTVHSLNRSPEPLVAEIVLVDFSDREHLKKPLD--YMAFPSPVRLITKKRGLI 205
Qy      219 RTMGLASVATGVDVITFLDSHCANVWMLPPLDIRIARNKRTIVCPMIDVIDHDFRYET 278
Db      206 RSRVRGADAAQAKVLTFLDSHCECHHEMLERVAEDRTRVVSPITDIVIMDNFQYVG 265
Qy      279 QADQAMKATFDMWYKRIPIPELQKA---DPSDPESPVMAGGLFANDRKWMEIGY 335
Db      266 ASAD-LKGFDFWMLVFWMYTPPEQRSRQGNPAPIKTPMIAAGLFVMDKFYEBELGY 324
Qy      336 DPGLEIWGEBOYEISFKVMCGGRMEDIPCSRVGHIYKRYVYKVP--AGVSLARNLKRV 393
Db      325 DMMMDVWGENLEISFVWQCGSLLEIIFCSRVGHVFRKQHPYTFPGSGTVPARNTKRA 384
Qy      394 AEVWMDYAEIYQRPPEYRLASADVAVOCKLRSSLNCKSPKFMPTKIAMDLPKFPYVP 453
Db      385 AEVWMDYAEIYQRPPEYRLASADVAVOCKLRSSLNCKSPKFMPTKIAMDLPKFPYVP 437
Qy      454 EPPAAMGEIRNVGTGICADTGHGALGSPRLBECVRGRGEAAMNMQVFTFTRREDIRP 513
Db      438 RVPDH-----QDIAGAL----- 450
Qy      514 GDPOHTKKCFDPAISHTSP--VTLYDCHSMKGNOLMKRDKOKLYH-----PVSGS 562
Db      451 --QOQGT--CLDTIGHFADGVGVYEGHAGAGNOMALTREKSVKAMDLCITVDRAPGS 506
Qy      563 CM---DCSESDHR-----TFMNTCPSSLTQOMLP 589
Db      507 LTKLGGRENDSRQKWEIGNSKLRHVGSNLCIDSRATKSGLSVEVCGP-ALSQQWKF 565

RESULT 13
T42248
polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T42248; T23138
R:Hagen, F.K.; Nehrke, K.
J. Biol. Chem. 273, 8268-8277, 1998
A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:polypeptide
A:Reference number: 222126; MUID:98192620; PMID:9525933
A:Accession: T42248
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-618 <HAG>
A:Cross-references: UNIPROT:O61394; EMBL:AF031838; NID:g3047296; PIDN:AA033674.1; PID:g3
R:Barlow, K.
submitted to the EMBL Data Library, June 1998
A:Reference number: 219695
A:Accession: T23138
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-618 <WTL>
A:Cross-references: EMBL:AL024499; PIDN:CAA19707.1; GSPDB:GN00021; CESP:H38K22.5
A:Experimental source: Clone H38K22
A:Gene: H38K22.5
A:Map position: 3
A:Introns: 42/3; 64/3; 154/1; 202/1; 368/3; 430/1; 477/3; 513/3; 563/3; 604/3
A>Note: gly-6

```

```

C:Superfamily: polypeptide N-acetylglucosaminyltransferase
C:Keywords: glycosyltransferase; hexosyltransferase
Query Match      28.9%; Score 948.5; DB 2; Length 618;
Best Local Similarity 38.3%; Pred. No. 1.3e-68;
Matches 211; Conservative 100; Mismatches 163; Indels 77; Gaps 21;

Qy      73 DMDHKAIRDAQVNGEGQRPYPMTDARV--DQYRNGFNIVYSDKISLNRSLDIR 131
Db      97 DW-----GGGAGVSHLTPEQQLADSTPAVQFNLVSDGISTVRSLEIR 143
Qy      132 HPNCNSKRYLETLPNTSIIIPFNEGMSLLRTVHSLNRSPEPLVAEIVLVDFSDREH 191
Db      144 KPSGRNMTYEDNLPPTSVIIVYHNKASTLIRTWASVINDSPKLEKEIILVDFSDREF 203
Qy      192 LKKP-LEDYALPFS-VRIITKTKREGIITRMGLASVATGVDVITFLDSHCANVWMLP 249
Db      204 LRYPTLDTTLKPLPTDIKTIIRSKERVGLIRARWGAQEAQGDVLTFLDSHCECTKWLEP 263
Qy      250 LDIRIARNKRTIVCPMIDVIDHDFRYETQAGDAMKATFDMWYKRIPIPELQKA---A 306
Db      264 LRTIKLRRAVPCPVIDIINDNTFOYQ-KGIEFRGFGFWNNQFRWYGMETPAKQHL 322
Qy      307 DPSDPESPVMAGGLFANDRKWMEIGYDPGLEIWGEBOYEISFKVMCGGRMEDIPCS 366
Db      323 DPTPTISPTMAGGLFIRNNRYFEBELIYDPMIDWGENLEMSFRIMQCGRVEILPCS 382
Qy      367 RVGHIYKRYVYKVP--AGVSLARNLKRAVWMDYAEIYQRPPEYRLASADVAVO 422
Db      383 HGVHVFRRKSPHPDPGKSGKANTNLRLVAEVMWMDMKHYFYXIAQAHMRSSIDVSE 442
Qy      423 QKLRSSLNCKSPKFMPTKIAMDLPKFPYVPEPPAAMGEI-RNVGICLADTGHGALGS 481
Db      443 RVELRKKLNCKSPKFMPTKIAMDLPKFPYVPEPPAAMGEI-RNVGICLADTGHGALGS 490
Qy      482 PLRLEG-----CVARGBAAMNMQVFTFTRREDIRPQDPOHTKKCFDAIS--HTS 531
Db      491 TLNSSGKITKTSYADCLK-----IFHKTQMLWYTGDRIRIRIDE-----HLCLSVQLAHTT 540
Qy      532 ---PVTLYDCHSMKGNOLMKYR-KDKTLYHPVSGSCMDCESDHRIP-----MN 576
Db      541 SDWKIQKEQAGF-DTEYWFKPKRIGRFQNRKTKGLCLASPD---IFDPTDBEFNPPIVQ 595
Qy      577 TCNPSLTQOM 587
Db      596 KCRSSNDQRNW 606

RESULT 14
T42250
polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) 6c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T42250
R:Hagen, F.K.; Nehrke, K.
J. Biol. Chem. 273, 8268-8277, 1998
A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:polypeptide
A:Reference number: 222126; MUID:98192620; PMID:9525933
A:Accession: T42250
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-562 <HAG>
A:Cross-references: UNIPROT:O61396; EMBL:AF031840; NID:g3047200; PIDN:AA033676.1; PID:g3
A:Gene: gly-6
C:Superfamily: polypeptide N-acetylglucosaminyltransferase
C:Keywords: glycosyltransferase; hexosyltransferase
Query Match      28.6%; Score 938; DB 2; Length 562;
Best Local Similarity 47.5%; Pred. No. 8.1e-68;
Matches 182; Conservative 74; Mismatches 103; Indels 24; Gaps 8;

Qy      73 DMDHKAIRDAQVNGEGQRPYPMTDARV--DQYRNGFNIVYSDKISLNRSLDIR 131

```

```

Db 97 -----GEGAGVSHLTPEQQLADSTFVAVNQFNILVSDGISVRSLEIR 143
Qy 132 HPNNSKRYLETLENTSIIIPFNHEGWSLLRTVSHVLRSPPELVAVIYVNDPSPREH 191
Db 144 KPSERNNTYIPDNLPTTSVITVYHNEAVSTLLRTVSWVIDRSPKELKEIILVDDPSREF 203
Qy 192 LKRC-LBDYMA/LPFS-VYIARTKKREGILRTFRLMGASVATGDIYTFDLSHCEAVNM/LP 249
Db 204 LRYTLDTLTKPLPTDILKIRSKERVGLIRARMMAGAGQGDVITFLDSSHECTKGMLEP 263
Qy 250 LLDRIANRRTIVCEPMIDVIDHDDFRYETQAGDAMRGAFDWMYKRIPIPELOK---A 306
Db 264 LLTRIKLNKRAKVPCEPVIDIINDNTFOYQ-KGIEHMRGGFNWNLQFRWYGMPTAMAKQILL 322
Qy 307 DPSDPFSPVWAGGLFAVDRKMFELGSDYDGLIHWGEQYEISFKVMCCGRMEDIPCS 366
Db 323 DPTGPISPTWAGGLFISINRYFELGSDYDGMIDIMGEMLEMSFRLWQCGRVEILPFS 382
Qy 367 RVGIYRYKVPYKVP---AGVSLARNLKRVAVMMDEYAYIYORRPE-YRHLASGVAV 422
Db 383 HVGIVFRKSSPHDPCKSGKVLNTNLLRVAEVMMDDMKHYFKIARQAHMRSSIDVSE 442
Qy 423 QKKLRSSLNCKSPFKWMTKIAMD 445
Db 443 RVELRKKLNCKSPFKWYLONVFOD 465

```

## RESULT 15

```

JC5247
poly(amide N-acetyl)galactosaminyltransferase (EC 2.4.1.41) T3 - mouse
N/Alternate names: protein-UDP acetyl)galactosaminyltransferase
C/Species: Mus musculus (house mouse)
C/Date: 13-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C/Accession: J05247
C/Author: J. K. Ten Hagen, F. K. Ten Hagen, K. G. Van Whyckhuysse, B. C. Tabak, L. A.
Biochem. Biophys. Res. Commun. 228, 38-44, 1996
A/Title: Cloning and expression of mouse UDP-GalNAc:poly(amide N-acetyl)galactosaminyltr
A/Reference number: J05247; MUID:97069650; PMID:8912633
A/Accession: J05247
A/Molecule type: mRNA
A/Residues: 1-633 <ZAR>
A/Cross-references: UNIPROT:P70419; GB:U70538; NID:91575722; PIDN:AA09579.1; PID:915757
A/Experimental source: testis
C/Comment: This enzyme catalyzes the formation of N-acetyl-D-galactosamine alpha-O-Ser/T
ictating O-glycosylation of serine and threonine residues on an array of glycoproteins.
C/Superfamily: poly(amide N-acetyl)galactosaminyltransferase
C/Keywords: glycosyltransferase; hexosyltransferase

```

Query Match 28.5%; Score 934; DB 2; Length 633;

Best Local Similarity 39.2%; Pred. No. 2e-67;

Matches 206; Conservative 98; Mismatches 179; Indels 42; Gaps 18;

```

Qy 90 GEGRPYPMTDARVDAQYRENG---FNIVSDKISLNSL-PDIRHPNCSKRY--L 141
Db 122 GASGKPFKITHLSPBEQKERGETKQCFNAPASDRISLHRDLGDTAPPCIEQKFKRC 181
Qy 142 ETLNTEGIIIPFNHEGWSLLRTVSHVLRSPPELVAVIYVNDPSPREH/LKPEYMA 201
Db 182 PPLPTTSVITVYHNEAVSTLLRTVSWVIDRSPKELKEIILVDDASVDYDLHEKLEEK 241
Qy 202 LFPSEVRIARTKKREGILRTFRLMGASVATGDIYTFDLSHCEAVNM/LP/LRIANRRTI 261
Db 242 QFSIVKIVRQGERKGLIARLILGAVAFAETLTFDADCEGCEYGLLEPLARIENYAV 301
Qy 262 VCEPMIDVIDHDDFRYETQAGDAMRGAFDWMYKRIPIPELOKADPSDFSPV 316
Db 302 VSPDIASIDLNTFERNKSPYGSNNRGNFWSLSFGWESL/PDHEKQKRKDEYPIKPT 361
Qy 317 MAGGLFAVDRKMFELGSDYDGLIHWGEQYEISFKVMCCGRMEDIPCSVGHYIRKV 376
Db 362 FAGGLFISIKYFEHIGSYDEEMETWGBENIEMSPRWQCGGLIMPCSVVGHVFRKS 421

```

```

Qy 377 PYKVPAGVS-LARNLKRVAEVMDEYAYIYORRPE---YRHLASGVAVQKKLRSSLN 431
Db 422 PHTEPKGTQYIARNOVALAEVMMDEYKIEFYRRTDAAKIVQKSPFSLSRFEIKRLO 481
Qy 432 CKSFKWMTKIAMDLPEFYRPEVPPAAANGEIRVVGGLCADT-KHGALSGPLRLECYR 490
Db 482 CKNFTWYLNIT---YPRAYVDPNLNVIS-GYKSVGPPCLDVGENNQGGKPLLYTC-H 536
Qy 491 GRGEAANNMNOVFETFRREDIRPDPOHT--KKFCPDASHTSPV--TLYDCS--MKGN 544
Db 537 GLG-----GNQYFYSAGQREIR-----HNQKELCHLHATGQVGVQKACVYKGHRTIAPG 586
Qy 545 QLMRYRDKTLVHEVSGSCMDCSSEDRIFMNTCNPSLTQQLF 589
Db 587 QIVEIRDOQLYNLPLFKVCLIS-SWGBHPNLY-PCDADTLQKWIIF 629

```

## RESULT 16

T31549

poly(amide N-acetyl)galactosaminyltransferase (EC 2.4.1.41) 9 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004

C/Accession: T31549; T42253

R/Matthews, L.

submitted to the EMBL Data Library, October 1999

A/Reference number: 221043

A/Accession: T31549

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-579 <WIL>

A/Cross-references: UNIPROT:Q9U2C4; EMBL:AL117202; PIDN:CA057897.1; CESP:Y47D3A.23

A/Experimental source: clone Y47D3A

R/Hagen, F. K.; Nehrke, K.

J. Biol. Chem. 273, 8268-8277, 1998

A/Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine:poly(pect

A/Reference number: 222126; MUID:98192620; PMID:9525933

A/Accession: T42253

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-202; 'D', 204-579 <HAG>

A/Cross-references: EMBL:AF031843; NID:93047206; PIDN:AA033679.1; PID:93047207

C/Genetics:

A/Gene: CESP:Y47D3A.23; gly-9

A/Introns: 45/1; 179/1; 343/3; 373/1; 453/3; 491/3; 533/3

C/Superfamily: poly(amide N-acetyl)galactosaminyltransferase

C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 27.6%; Score 906; DB 2; Length 579;

Best Local Similarity 40.2%; Pred. No. 3.3e-65;

Matches 212; Conservative 88; Mismatches 182; Indels 46; Gaps 22;

```

Qy 84 AQRVNGEGGRPYMT--DARVDAQYRENG-FNIVSDKISLNSL-PDIRHPNCSKRY 140
Db 70 APRGPGKGPVVLVTGDAE-LQADMKKFMVNHASDKLSLDVDFDPRIQAKCKDITK 128
Qy 141 -LETLNTEGIIIPFNHEGWSLLRTVSHVLRSPPELVAVIYVNDPSPREH/LKPEYMA 199
Db 129 DYALPPTSIVIIITDEAMTFLRTVSHVLRSPPELVAVIYVNDPSPREH/LKPEYMA 188
Qy 200 MALP-PSYRIARTKKREGILRTFRLMGASVATGDIYTFDLSHCEAVNM/LP/LRIANR 258
Db 189 IKRGGKVRLLRKVRHGLIPAKLAGAEVAGDIIIVFLDSCENHGLLEPIVORISDER 248
Qy 259 KTIVCPMIDVIDHDDFRYETQAGD-AMRGAFDWMYKRIPIPELOK--ADPSDFSP 314
Db 249 TALVCEPMIDISDNTLAH--GDWSLSTGGSFSAALHTWEGLSBEEQKRTKPTDYRS 305
Qy 315 PVNAGGLFAVDRKMFELGSDYDGLIHWGEQYEISFKVMCCGRMEDIPCSVGHYIRKV 374
Db 306 PTMAGGLAANREKFEVVGVDSEMDIHWGEMLEMSFRLWQCGGLIMPCSVVGHVFRKS 365
Qy 375 VYKVPAGVS---LARNLKRVAEVMDEYAYIYORRPEYRHLASGVAVQKKLRSSL 430

```



Db 366 GHPYNTGRNNKOVHGTNSKRLAEVMMDDYKRLYYVHREDLRTKOVGDLTARHBLKRL 425  
 Qy 431 NCSSEFMFKIMDKLPKFPVPPPAANGELRN--GTGLCADT--KRGALGSPRLKE 466  
 Db 426 NCRPFKMFIDNIA--KCKFI--MDQVAVAGALHTVSGTRMCTDILQDERKNSQLGVP 481  
 Qy 487 GCVRGGEAAMNQVFTPTWRBDIRPGDQHTKKFCFDIAISHTSPVTLTDCSHMKG--N 544  
 Db 482 HC--QGGK---SSPOLMSLSEKGNLR-----RENTC--ASEENGNIIRMTTC--SKKAQFN 527  
 Qy 545 QLMKRYKDKTLHPVSGSCMDCE---SDHRIFMNTCNPSSLTQOMLF 589  
 Db 528 ERWAY--ENKMIIRNLKSKCKMSTANLTKGDAIYVE--CDEDEHQKNMF 573

## RESULT 17

T42244

probable polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C/Accession: T42244  
 J:Hagen, F.K.; Nehrke, K.  
 J. Biol. Chem. 273, 8268-8277, 1998  
 A>Title: cDNA cloning and expression of a family of UDP-N-acetyl-Dgalactosamine:Polypept  
 A/Reference number: Z22126; MUID:98192620; PMID:9525933  
 A/Accession: T42244  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-589 <HAG>  
 A/Cross-references: UNIPROT:O61390; EMBL:AF031834; NID:G3047188; PIDN:AA013670.1; PID:G3  
 C:Genetics:  
 A:Gene: gly-4  
 C:Superfamily: polypeptide N-acetylglactosaminyltransferase  
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 26.9%; Score 883; DB 2; Length 589;  
 Best Local Similarity 37.4%; Pred. No. 2.5e-63;  
 Matches 209; Conservative 80; Mismatches 188; Indels 82; Gaps 18;

Qy 56 HSRQK-----KTF---FLDGQGLKDMHDKAIRDADQVNGEGQRPYMTDAERVDQ 106  
 Db 83 HERTKDVTKTDFVEKFLNKGK---WH-----QG-----ED 111  
 Qy 107 AYRENGFNIVSDKISLNSLPDIRHPNCSKRYLET--LENTSIIIPFNHSGSLIKTY 165  
 Db 112 KYRANFENOBASDALNPTRKIPDSREPOCRDVYSKVGQPTVITIIYHNEASSLIKTY 171  
 Qy 166 HSYLANSPELVVAEIVLVDDFSDREHLKKPLEDYMAFPESVRIILRTKKRGLRTMYLGA 225  
 Db 172 FSYFNOSPEELAEIVLVDDNSQDVEIGKEL---AQIQRIIVLRNNQREGILRSRVKGA 227  
 Qy 226 SVATGVDITFLDSHCANVMPLPLDRIARNRKTYICPMIDVIDHDHDFEYEQADAMR 285  
 Db 228 QVABAPVLTFLDSHIECNQKMLBELIARIENPRAVVAPIIIDVINDFNYGASAD--LR 286  
 Qy 286 GADWEMTYRKIRIPELQK---ADPSDPFESPMAGGLFAVDRKFWELGGIDPGLEIV 342  
 Db 287 GGDWMLTVFWEEMNEQLRKERHAHPAPYRSPMAGGLFALSKEMFNEIGTVDLMEVW 346  
 Qy 343 GGEQVEISFYVMWCGGEMEDIPCSRVGHIYRKYVYVYVPAVGS--LABNLKRAVAEVMDE 400  
 Db 347 GGNLEISFVWQGGSLLEIPCSRGVHFRKGIPTFPFGSGNVFOKNTRRRAAEVWDE 406  
 Qy 401 YAEIYORREYRHLISAGDVAVQKGLASSLNCKSPKMFMTKIAMDLPEKFPYVPEPAAW 460  
 Db 407 YKAIYIKNVBSARFVNVGDIITDRAIRDLQCKSPKMYLNV-----YPLEIRPKTP 459  
 Qy 461 GELRWGTC--LCADYTKAGALGSPRLKEGCVRGGEAAMNQVF---TFTWRREDIRGDP 516  
 Db 460 GKSPQKIGTICLDSPARKSEAPGLFGCHGTGNGDEM---VEDQTLTKFKNAI----- 510  
 Qy 517 QHTKKCFDAISHT--SPVTLTDCSHMKGNQMLKRYRDKTLTLPVSGSCMDCE---SDH 571

Db 511 ---SGLCLDPSNTENKTVTWKCENTLRPDTM---VVEKNQMLTGKGCITLVNQSGGDW 564  
 Qy 572 RIFMNTCNPSSLTQOMLF 590  
 Db 565 LIYGACELNNGAQRIWFE 583

## RESULT 18

T27397

hypothetical protein Y75B8A.9 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C/Accession: T27397  
 J:Barlow, K.  
 Submitted to the EMBL Data Library, November 1998  
 A/Reference number: Z20361  
 A/Accession: T27397  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-605 <WIL>  
 A/Cross-references: UNIPROT:Q9XW72; EMBL:AL033514; PIDN:CAA22098.1; CBSP:Y75B8A.9  
 C:Genetics:  
 A:Gene: CRSP:Y75B8A.9  
 A:Interons: 78/3; 251/3; 391/2; 529/3; 565/3  
 C:Superfamily: polypeptide N-acetylglactosaminyltransferase

Query Match 24.4%; Score 801; DB 2; Length 605;  
 Best Local Similarity 31.9%; Pred. No. 1.1e-56;  
 Matches 187; Conservative 102; Mismatches 194; Indels 103; Gaps 18;

Qy 80 IRDDAQVNGEGGR-----PYMTDAERVD----- 105  
 Db 44 IORHEKLNNGQGRHDPDDDEGAKEQBEDAVEKONIAAPPLKSFPTPDRSKELIDT 103  
 Qy 106 -----QAYRENGFNIVSDKISLNSLPDIRHPNCSKRYLETLPNTSI 149  
 Db 104 DLGKINGRAEDLQVGYKYQFNGLSDRISKIDRSRARSLSLTSLSPLAASI 163  
 Qy 150 IIFPNHSGSLIKTYHVSVLNSRSPPELVVAEIVLVDDFSD---REHLKKPLEDYMALFPS 205  
 Db 164 VVCYFNBPSEVLIRVNSIFDRTPKPHLHILLVDDSSSEWSNATDEAIKYREKHIIQWEK 223  
 Qy 206 VRLIRTKKREGLIRTMGLASVATGDIYITFLDSHCANVMPLPLDRIARNRKTYICPM 265  
 Db 224 VKELKTDKNEGLIRAKIFGARBRANGEVLVFLDSHCENBEWELPPLDQIKQNRARRVCP 283  
 Qy 266 IDVIDDDPRYEQADAMRGAFDWEYTKRIIPELQKADS---DPFESPMAGGLF 322  
 Db 284 IDIIDLITMKY--VESPVCTGGVNMAMTRK--WDYFHSYFEDMNTVNPILKSPMAGGLF 340  
 Qy 323 AVDRKFWELGSDYDPSGLEIWSGEQVEISFYVMWCGGEMEDIPCSRVGHIYRKYVYVPA 382  
 Db 341 AIDKEVFEELIGSDDEQMDVWGAENVEISVRIWCGGELLIMPESRGHIFRQRPQIKT 400  
 Qy 383 GVSLLANLKRVAEVMWDEIAEYIYORREYRHLIS--AGDVAVQKGLASSLNCKSPKMFMTK 441  
 Db 401 D--SMGRNVSRLAVWVWDEIYENFEARPNRYPTDGLDLSRLSLRNLOCKPEKYLEN 459  
 Qy 442 IAMDLPKFPVPPPAANGELRNV-----GTG--LCADYTKG--ALGSPRL 484  
 Db 460 I---YBELLDNTP-----NQLNNQILVAGKTLIMANGTCHLSBNSGRLANGNVE 511  
 Qy 485 LBSGVRGGEAAMNQVFTWRREDIRPDPQHTKKFCFDIAISHTSPVTLTDCSHMKGN 544  
 Db 512 MRCKNH-----MERNQOMKYSTNELR---PWGSSRMCLDSLRGIVL---CHNGAH 559  
 Qy 545 QLMKRYKDKTLHPVSGSCMDCESDHRI--FNATCNPSSLTQOMLF 589  
 Db 560 QMWQVSNAGRLY---SRVYKCATGSDNVSAISTLKFCSLANSFOF 602

## RESULT 19

T42252  
poly(amide N-acetylglucosaminyl)transferase (EC 2.4.1.41) 8 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T42252; T27290  
R:Hagen, F.K.; Nehrke, K.  
J. Biol. Chem. 273, 8268-8277, 1998  
A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:poly(amide N-acetylglucosaminyl) transferase  
A:Reference number: 222126; MUID:98192620; PMID:9525933  
A:Accession: T42252  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-421 <HAG>  
A:Cross-references: UNIPROT:O45293; EMBL:AF031842; NID:G3047204; PIDN:AA013678.1; PID:G3047204  
R:Matthews, L.  
submitted to the EMBL Data Library, October 1998  
A:Reference number: 220338  
A:Accession: T27290  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-421 <MTL>  
A:Cross-references: EMBL:AL032622; PIDN:CAA21500.1; GSPDB:GN00021; CESP:Y66A7A.6  
A:Experimental source: clone Y66A7A  
C:Genetics:  
A:Gene: gly-8; Y66A7A.6  
A:Map position: 3  
A:Insertion: 36/2; 62/3; 97/2; 181/3; 226/2; 279/2; 318/3; 369/3  
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 19.3%; Score 632; DB 2; Length 421;  
Best Local Similarity 34.0%; Pred. No. 3.4e-43;  
Matches 138; Conservative 78; Mismatches 146; Indels 44; Gaps 10;

QY 83 DAGVNGGEQGRPY-----PMIDAEVDAQYRN-----GFIY 116  
DB 21 EAEGLPKCEHVDPEYENEGMIDLPKLT-RKCNHTLKENLTBAASKSEMGKSPAPDAL 79  
QY 117 VSDKISLNRSLPDRHPNCSKRYLETLPNTSIIIPFNEGMSLLRTSHSVLRSPPEL 176  
DB 80 SSELGPRNRYGKQAHKCEEEK-DASVTSVIVHNEALSTILNINIGIETPRLSL 138  
QY 177 VAEIVLVDDPSDRH-LKKPLEDY---MALPSPVRIITFKREGILRTMIGASVATGDV 232  
DB 139 LKEIVLVDDPSDRH-LKKPLEDY---MALPSPVRIITFKREGILRTMIGASVATGDV 198  
QY 233 ITPLDSHCEANVMPLPDLIRARNRKTIVCPMIDVIDHDHFRYETQAGMARGAFDEM 292  
DB 199 IVEFDSHCEANVMPLPDLIRARNRKTIVCPMIDVIDHDHFRYETQAGMARGAFDEM 256  
QY 293 YKRRIPIPELQKADPSD--PFESPVNAGGLFAVDKRMFELGVDPLGELTIGSEQYEIS 350  
DB 257 TFKKIYLPWEYETFPENNVRKFNPSAMPGGLAMRKCFVFLGDEMDMGEISENIELS 316  
QY 351 FKVMCGRMEDIPCSRVRGHTYRKVPYKVPAGVSLA-RNLKRVAEVMDYAEIYQRR 409  
DB 317 LKAMICGGRVVAACSRGVHFRMRPYTSRKGMDALYNNAVRAKTMGLGESKFAVK 376  
QY 410 PEYHLSAGDVAVQKURSSLNCKSEFKEMTKIAMDLKPYPPVER 455  
DB 377 PRGAKMVFGLTPEMVOVKRLCKDMKFIENY-----YPELEP 415

RESULT 20  
H97186  
glycosyltransferase domain containing protein [imported] - *Clostridium acetobutylicum*  
C:Species: *Clostridium acetobutylicum*  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: H97186  
R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum* ATCC824  
A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: H97186  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1044 <KUR>  
A:Cross-references: UNIPROT:Q97GN9; GB:AE001437; PIDN:AAK80283.1; PID:G15025335; GSPDB:G15025335  
A:Experimental source: *Clostridium acetobutylicum* ATCC824  
C:Genetics:  
A:Gene: CAC2327

Query Match 4.5%; Score 147; DB 2; Length 1044;  
Best Local Similarity 20.6%; Pred. No. 0.0025;  
Matches 48; Conservative 55; Mismatches 108; Indels 22; Gaps 8;

QY 141 LETLPTSIIPFNFNEGMSLLRTSHSVLRSPPELVAVIVLVDPS---DREHLKKPLR 197  
DB 1 MNSVPKVSFIIVVNN-GLQHLKNCFSLSKLSYSYSDKIEIVLVNDGSKDSVETLKN-- 57  
QY 198 DYMLPSPVRIITFKREGILRTMIGASVATGDVITFLDSHCEANVMPLPDLIRAR- 256  
DB 58 ----YPAVVIITKDSNEGFAPKPDPAKTAEGEYALINNDMLDKNWNIDMEETLENC 112  
QY 257 NRKTIVCPMIDVIDHDHFRYETQAGMARGAFDEMYYKRIPIPELQKADPSDPFESPV 316  
DB 113 NDNSIVCAGSKIVWDSKIDPAGSVSPAGYGIQYDGM----DIKANKKTNEBDRI 167  
QY 317 M--AGGLFAVDKRMFELGVDPLGELTIGSEQYEISFKVMCGRMEDIPCSR 367  
DB 168 LFAAGSMILRKDVFIETIGGFDKDYFAV-YEDVDLGRMLVGLYKFAV--CSK 217

RESULT 21  
D87531  
glycosyl transferase family protein CC2277 [imported] - *Caulobacter crescentus*  
C:Species: *Caulobacter crescentus*  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C:Accession: D87531  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.; B.; Taub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonitskii, J.; Brumlaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: D87531  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-306 <STO>  
A:Cross-references: UNIPROT:Q9A618; GB:AE005673; NID:G13423792; PIDN:AAK24248.1; GSPDB:G13423792  
C:Superfamily: Rhizobium succinoglycan biosynthesis glycosyltransferase

Query Match 4.2%; Score 139; DB 2; Length 306;  
Best Local Similarity 25.4%; Pred. No. 0.0019;  
Matches 57; Conservative 36; Mismatches 91; Indels 40; Gaps 10;

QY 144 LPTSIIPFNFNFNEGMSLLRTSHSVLRSPPELVAVIVLVDPSDRHLKKPLEDYMA 201  
DB 1 MARTVMIIPQGRPDGLAAVARSFVGVDPAEL--ELVIYD--NQVSAKAVADALR 56  
QY 202 L-PPSVRIITFKREGILRTMIGASVATGDVITFLDSHCEANVMPLPDLIRARNKT 260  
DB 57 KGAPCPVIYVNEKRGVAFARNAGMARASGFIIFLDDDEBAPSGWLAALAAOERYDAD 116  
QY 261 IYCPMI-----DIVDDDF--RYETQAGMARGAFDEM--EMYYKRIPIPELQKAD 307  
DB 117 VFEFVAPARAPAHIDQRDIERFSGRIQDAQVIVHYGCGSLIRSLP-----D 170  
QY 308 PSDP-----ESPVMAGGLFAVDKRMFELGVDPLGELTIG 342  
DB 171 PVAFAVARNFISGEDDLRGMGAAGKRAWE-----PAAMW 209

RESULT 22





**RESULT 25**

D87506

glycosyl transferase family protein CC2077 [Imported] - *Caulobacter crescentus*

C:\Date: 20-Apr-2001 #sequence\_ revision 20-Apr-2001 #text\_change 09-Jul-2004

C;Accession: D87506

1. The first group of respondents (n = 10) was composed of students who had completed the course and were currently employed in a related field. 2. The second group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 3. The third group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 4. The fourth group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 5. The fifth group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 6. The sixth group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 7. The seventh group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 8. The eighth group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 9. The ninth group (n = 10) was composed of students who had completed the course and were currently employed in a related field. 10. The tenth group (n = 10) was composed of students who had completed the course and were currently employed in a related field.

C;Accession: D87506

R;Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hatt, D.H.; Kolon, J.; Ermojevna M.; White O.; Salzberg S.L.; Shavrylo I.; Yenter J.C.; Eraser C M.

Proc. Natl. Acad. Sci. U.S.A. 98: 4136-4141, 2001

A:Title: Complete Genome Sequence of *Caulobacter crescentus*.  
FIOC: NAL1:ACAU: SCI: 0:3:A: 30, 4138-4141, 2001

A:Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: D87506

A;Status: preliminary

A;Molecule type: DNA

```
A;Residues: 1-318 <STO>
```

A;CROSS-REFERENCES: UNL  
C:Genetjca:

C:\GENELICS:  
A:Gene: CC2077

11/03/2011 11:03:00

Query Match

Best Local Similarity 24.7%; Pred. No. 0.0039;

Matches 63; Conservative 39; Mismatches 102; Indels 51; Gaps 10;

143 ТИ ДНТЕСТТТДЕНУЕСИУСИ Т ДТНУЕУТ НРБЕРБЕИ УАЕТИТ УНДЕСДРБЕУ ККДТ ЕДУМАТ. 202

143 TLENTSILIPFNEGWSLLRIVHSV LNRSPPELVAEI LVDDFS DREHLKKPLEDMAL 202

```
Db      |   ||| |  
       |   ||| |  
17 TRPRSLVMVYMTG-PALMESIRHALDE---PRVDEFITVDNGSSILADAAW-LRDIARR 71
```

[illegible]

203 FPSVRLRTKKREGILRTPLGASVATGDVITFLDSHC EANVNLPP---LDRIARNR 258

Db 72 EPRVRLQGLNIGFARAANLGATAKGEDEVFLNPDA-----FLTPGAIAALREARDR 126

200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000

259 K I I V C P M D V I D H D D F R Y E T Q A G D A M R G A F D ----- W E M Y Y K R L P 298

Db 127 P S P C V V G A R V E N T D G --- T E O R G G R G E I T P V T T I L S I S K I S A T L P L R R F E I H R E G E P 1822

127 F0FCV68A VENNIDG - - - 1E9K00K0B011FV111B00D0L0B011FF1NNF211NE0EF 102

Search completed: November 22, 2004, 13:44:12  
Job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 13:38:22 ; Search time 67 Seconds

(without alignments)  
5178.372 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278  
Sequence: 1 MRKEKRLQALVALVLAALV.....TQGMLEPHNTSYLAEKRN 603

Scoring table: BLOSUM62

Gapop 10.0 , Gapekt 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3278	100.0	603	1	GLI0_HUMAN
2	3157	96.3	603	1	GLI0_RAT
3	3138	95.7	634	2	BAD21A05
4	3137	95.7	603	1	GLI0_MOUSE
5	3137	95.7	603	2	AAH60617
6	1404	42.8	644	2	0700B9
7	1392.5	42.5	622	1	GLI0_CAEEL
8	1360	41.5	666	1	GLT6_DROME
9	1360	41.5	666	2	AAQ56703
10	1341.5	40.9	599	2	07PZM5
11	1247	38.0	1003	2	070DRO
12	1186	36.2	659	1	GLT4_DROME
13	1186	36.2	659	2	AAQ56701
14	1172.5	35.8	650	1	GLT9_DROME
15	1159	35.4	645	2	07Q815
16	1144	34.9	518	2	07Q048
17	1125	34.3	559	1	GLT1_HUMAN
18	1117	34.1	559	1	GLT1_PIG
19	1116	34.0	559	1	GLT1_RAT
20	1115	34.0	559	1	GLT1_BOVIN
21	1115	34.0	559	1	GLT1_MOUSE
22	1113	34.0	601	1	GLT7_CAEEL
23	1112.5	33.9	556	2	06GM51
24	1111.5	33.9	617	2	AAQ56702
25	1109	33.8	556	1	GLI3_MOUSE
26	1109	33.8	556	1	GLI3_RAT
27	1109	33.8	556	1	AAQ75749
28	1108.5	33.8	630	1	GLT5_DROME
29	1106	33.7	556	1	GLI3_HUMAN
30	1087	33.2	559	2	06PAB1
31	1087	33.2	559	2	AAH60419

32	1083	33.0	626	1	GLT5_CAEEL	095211 caenorhabdi
33	1083	33.0	653	2	06NS24	06ns24 xenopus lae
34	1083	33.0	653	2	AAH70527	AAH70527 xenopus l
35	1053	32.1	653	2	06NX21	06nx21 xenopus tro
36	1053	32.1	653	2	AAH67317	AAH67317 xenopus t
37	1050	32.0	657	1	GLT7_HUMAN	086822 homo sapien
38	1049	32.0	930	1	GLT5_MOUSE	086102 m polypepti
39	1047	31.9	657	1	GLT7_MOUSE	080v30 mus musculu
40	1040	31.7	657	1	GLT7_RAT	097053 rattus norv
41	1040	31.7	930	1	GLT5_RAT	088422 r polypepti
42	1037	31.6	606	2	07QEH0	07qeh0 anopheles g
43	1036	31.6	578	1	GLT4_HUMAN	08n440 h polypepti
44	1035.5	31.6	591	1	GLT7_DROME	08mva8 drosophila
45	1035.5	31.6	591	2	AAQ56704	AAQ56704 drosophi
46	1035	31.6	582	2	06IR91	06ir91 xenopus lae
47	1035	31.6	582	2	AAH71009	AAH71009 xenopus l
48	1033	31.5	578	1	GLT4_MOUSE	008832 m polypepti
49	1033	31.5	578	2	AAH57882	AAH57882 mus muscu
50	1013	30.9	940	1	GLT5_HUMAN	07z769 h polypepti
51	999.5	30.5	558	1	GLT1_HUMAN	08n428 h putative
52	990.5	30.2	576	1	GLI2_MOUSE	08bgc3 m polypepti
53	990.5	30.2	563	2	AAH56425	AAH56425 mus muscu
54	989	30.2	563	2	06DJR8	06dj8 xenopus tro
55	988.5	30.2	581	1	GLT3_CAEEL	08ixk2 h polypepti
56	986.5	30.1	581	1	GLI2_HUMAN	08ixk2 h polypepti
57	977.5	29.8	558	1	GLT1_MOUSE	09j161 m putative
58	971	29.6	552	1	GLI4_HUMAN	096fi1 h polypepti
59	971	29.6	552	2	AAQ89118	AAQ89118 homo sapi
60	965	29.4	550	1	GLI4_MOUSE	08bvg5 m polypepti
61	964	29.4	620	2	07Q456	07q456 anopheles g
62	957.5	29.2	608	2	GLI1_MOUSE	092118 m polypepti
63	957.5	29.2	608	2	AAQ06668	AAQ06668 mus muscu
64	954.5	29.1	608	1	GLI1_HUMAN	08ncw6 h polypepti
65	954.5	29.1	608	1	GLI1_RAT	08p6v1 r polypepti
66	954.5	29.1	608	2	AAH62004	AAH62004 rattus no
67	953.5	29.1	590	2	AAH64620	AAH64620 drosophi
68	953.5	29.1	633	1	GLT2_DROME	06wv19 drosophila
69	949	29.0	571	1	GLT2_HUMAN	010471 h polypepti
70	948.5	28.9	618	1	GLT6_CAEEL	AA61394 caenorhabdi
71	947.5	28.9	615	2	AAQ56700	AAQ56700 drosophi
72	942.5	28.8	570	1	GLT2_MOUSE	08pb93 m polypepti
73	942.5	28.8	570	2	AAH59818	AAH59818 mus muscu
74	935	28.5	633	1	GLT3_MOUSE	070419 m polypepti
75	926	28.2	639	1	GLT2_HUMAN	08n311 h polypepti
76	926	28.2	639	2	AAQ88808	AAQ88808 homo sapi
77	925.5	28.2	622	1	GLT6_MOUSE	08c747 m polypepti
78	925	28.2	622	1	GLT6_HUMAN	08nc14 h polypepti
79	923.5	28.2	590	1	GLT8_DROME	09vut6 drosophila
80	920	28.1	667	1	GLT3_DROME	09y117 drosophila
81	916.5	28.0	601	2	AAQ56699	AAQ56699 drosophi
82	916	27.9	633	1	GLT3_HUMAN	014435 h polypepti
83	913.5	27.9	601	1	GLT1_DROME	06wv20 drosophila
84	906	27.6	579	1	GLT9_CAEEL	09uzc4 caenorhabdi
85	905	27.6	648	2	07QAS5	07qas5 anopheles g
86	898	27.4	630	1	GLI0_DROME	08ia43 drosophila
87	896	27.3	619	2	06NX01	06nx01 brachydanio
88	896	27.3	619	2	AAH67340	AAH67340 brachydan
89	896	27.3	627	2	07QIK5	07qik5 anopheles g
90	883	26.9	589	1	GLT4_CAEEL	08i136 caenorhabdi
91	873.5	26.6	638	1	GLT2_MOUSE	09d288 m polypepti
92	862	26.3	606	2	07Q046	07q046 anopheles g
93	818.5	25.0	598	1	GLT3_HUMAN	06is84 h putative
94	818.5	25.0	598	2	AAH67524	AAH67524 homo sapi
95	818.5	25.0	598	2	AAH67525	AAH67525 homo sapi
96	818.5	25.0	598	2	AAH69624	AAH69624 homo sapi
97	818.5	25.0	598	2	AAH69628	AAH69628 homo sapi
98	818.5	25.0	598	2	AAH69636	AAH69636 homo sapi
99	818.5	25.0	598	2	AAH69645	AAH69645 homo sapi
100	818.5	25.0	598	2	AAH69997	AAH69997 homo sapi

## ALIGNMENTS

RESULT 1  
 GL10\_HUMAN STANDARD, PRT, 603 AA.  
 AC Q86SR1; Q86V8; Q81XJ2; Q8TE2; Q961V2; Q9H8E1; Q9Y4M4;  
 DT 01-OCT-2004 (Rel. 45, Created)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DT 01-OCT-2004 (Rel. 45, Last sequence update)  
 DE Polypeptide N-acetylgalactosaminyltransferase 10 (EC 2.4.1.41)  
 DE (Protein-UDP acetylgalactosaminyltransferase 10) (UDP-  
 DE GalNAc:polypeptide N-acetylgalactosaminyltransferase 10) (Polypeptide  
 DE GalNAc transferase 10) (GalNAc-T10) (pp-GalNAse 10).  
 GN Name=ALNT10;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), ENZYME ACTIVITY, AND TISSUE  
 RP SPECIFICITY.  
 RC MEDLINE=22304871; PubMed=12417297;  
 RA Cheng L., Tachibana K., Zhang Y., Guo J.-M., Tachibana K.K.,  
 RA Kameyama A., Wang H., Hiruma T., Iwasaki H., Togayachi A., Kudo T.,  
 RA Narimatsu H.;  
 RT "Characterization of a novel human UDP-galNAc transferase, pp-galNAc-  
 RT T10.";  
 RL FEBS Lett. 531:115-121(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Bennett E.P.;  
 RL Submitted (AUG-2002) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 4).  
 RC TISSUE=Uterus;  
 RA Ansoer W., Winkner U., Mewes H.-W., Gaessenhuber J., Wiemann S.;  
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 5), AND SEQUENCE OF 191-603 FROM N.A.  
 RP (ISOFORM 1).  
 RC TISSUE=Kidney, Skin, and Testis;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stepien M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Muliyil S.J.,  
 RA Bork S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hilyk S.W.,  
 RA Villalón D.K., Wozny D.M., Sodergren E.O., Lu X., Gibbs R.A.,  
 RA Foley J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield A.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP SEQUENCE OF 6-366 FROM N.A. (ISOFORM 3), AND SEQUENCE OF 117-603 FROM  
 RP N.A. (ISOFORM 2).  
 RC TISSUE=Placenta, and Spleen;  
 RX PubMed=14702039; DOI=10.1038/ng1285;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Ohayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
 RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,

RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
 RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,  
 RA Yamazaki M., Nimomiya K., Ishibashi T., Yamashita H., Murakawa K.,  
 RA Fujimori K., Tanai H., Kimata M., Matanabe M., Hirao S., Chiba Y.,  
 RA Ishida S., Kanehori K., Takahashi S., Matanabe S., Yoshida M., Houchi T.,  
 RA Kusano Y., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
 RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,  
 RA Imose N., Muesashi K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,  
 RA Yoshihara Y., Matsunawa H., Ichihara T., Shiohara T., Sano S.,  
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki H., Matanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamazaki M., Matanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohnori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senda T.,  
 RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togashi T., Oyama M., Hata H., Matanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maeno Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,443 full-length human  
 RT cDNAs.";  
 RL Nat. Genet. 36:40-45(2004).  
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has activity toward Muc5Ac and E22 peptide  
 CC substrates.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium (By similarity).  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
 CC similarity).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named Isoforms=5;  
 CC Name=1;  
 CC IsoId=Q86SR1-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q86SR1-2; Sequence=VSP\_011209;  
 CC Note=No experimental confirmation available;  
 CC Name=3;  
 CC IsoId=Q86SR1-3; Sequence=VSP\_011212, VSP\_011213;  
 CC Note=No experimental confirmation available;  
 CC Name=4;  
 CC IsoId=Q86SR1-4; Sequence=VSP\_011207, VSP\_011208, VSP\_011214;  
 CC Note=No experimental confirmation available;  
 CC Name=5;  
 CC IsoId=Q86SR1-5; Sequence=VSP\_011210, VSP\_011211;  
 CC Note=No experimental confirmation available;  
 CC -1- TISSUE SPECIFICITY: Widely expressed. Expressed at high level in  
 CC small intestine, and at intermediate levels in stomach, pancreas,  
 CC ovary, thyroid gland and spleen. Weakly expressed in other  
 CC tissues.  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called Gnt motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC Gal/GalNAc-T motif), which is probably involved in catalytic  
 CC reaction and UDP-Gal binding (By similarity).  
 CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and  
 CC contributes to the glycopeptide specificity (By similarity).  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
 CC subfamily.  
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.  
 CC CAUTION: According to experiments made in rat, this enzyme is  
 CC unable to transfer GalNAc onto serine or threonine residue on the  
 CC protein receptor, but instead requires the prior addition of a  
 CC GalNAc on a peptide before adding additional GalNAc moieties,  
 CC thereby acting as a glycopeptide transferase.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; AB078145; BACS6890.1; -  
 DR EMBL; AJ050590; CAD44532.1; -  
 DR EMBL; AL096739; CAB46378.1; -  
 DR EMBL; BC007224; AAH07224.2; -  
 DR EMBL; BC050333; AAH50333.1; -  
 DR EMBL; BC072450; AAH72450.1; -  
 DR EMBL; AK023782; BAB14676.1; ALT-INT.  
 DR EMBL; AK074132; BAB84958.1; -  
 DR PIR; T12552; T12552.  
 DR HSSP; P26514; KKM.  
 DR Gene; HGNC:19873; GALNT10.  
 DR MIM; 608043; -  
 DR InterPro; IPR001173; Glyco\_trans\_2.  
 DR InterPro; IPR008997; RicinB-like.  
 DR InterPro; IPR000772; RicinB-lectin.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR Pfam; PF00652; RicinB-lectin; 3.  
 DR SMART; SM00458; RICIN\_1; LECTIN; 1.  
 DR PROSITE; PS50231; RICIN\_B-LECTIN; 1.  
 KW Alternative splicing; Calcium; Glycosyltransferase; Golgi stack;  
 Lectin; Manganese; Signal-anchor; Transferrase; Transmembrane.  
 FT DOMAIN 1 11 Cytoplasmic (Potential).  
 FT TRANSMEM 12 31 Signal-anchor for type II membrane  
 FT protein (Potential).  
 FT DOMAIN 32 603 Luminal (Potential).  
 FT DOMAIN 144 253 Catalytic subdomain A.  
 FT DOMAIN 311 373 Catalytic subdomain B.  
 FT DOMAIN 458 590 Ricin B-type lectin.  
 FT DISULFID 471 488 By similarity.  
 FT DISULFID 523 538 By similarity.  
 FT CARBOHYD 563 578 By similarity.  
 FT CARBOHYD 124 124 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 146 146 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 593 593 N-linked (GlcNAc... ) (Potential).  
 FT VARSPLIC 1 329 Missing (in isoform 4).  
 FT /FTID=VSP\_011207.  
 FT VARSPLIC 330 352 WELGYPGEIHWGEGEYISFK -> MLANRDELEAETS  
 FT /FTID=VSP\_011208.  
 FT VARSPLIC 190 251 Missing (in isoform 2).  
 FT /FTID=VSP\_011209.  
 FT VARSPLIC 190 202 EHLKRPLEDYVAL -> DLPTASTPSPVC (in  
 FT isoform 5).  
 FT VARSPLIC 203 603 /FTID=VSP\_011210.  
 FT /FTID=VSP\_011211.  
 FT VARSPLIC 354 366 WMCGRMEDTFC -> SGLSRKRYLGTA (in  
 FT isoform 3).  
 FT /FTID=VSP\_011212.

Query Match 100.0%; Score 3278; DB 1; Length 603;  
 Best Local Similarity 100.0%; Pred. No. 3e-257;  
 Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRKERLLQAVLVALLVLLPVNGLMALYREPODPTPGSGAVALPAAGQSHSRQK 60  
 DB 1 MRRKERLLQAVLVALLVLLPVNGLMALYREPODPTPGSGAVALPAAGQSHSRQK 60  
 QY 61 KTFPLGSGQKLKQWHDKEAIRRDQVNGNGQGRPYPTMTDERDOAYRENGFNIVYSDK 120  
 DB 61 KTFPLGSGQKLKQWHDKEAIRRDQVNGNGQGRPYPTMTDERDOAYRENGFNIVYSDK 120  
 QY 121 ISLNRSLPDIRHPNCKSKRYLETPTNTSIIPFNNEGSSLLRTVHSLVNRSPPELVAEI 180  
 DB 121 ISLNRSLPDIRHPNCKSKRYLETPTNTSIIPFNNEGSSLLRTVHSLVNRSPPELVAEI 180

QY 181 VYVDPDSRREHLKKPLEDYVALPPSVRIILPTKREGILFRMIGASVANGDVITFLDSHC 240  
 DB 181 VYVDPDSRREHLKKPLEDYVALPPSVRIILPTKREGILFRMIGASVANGDVITFLDSHC 240  
 QY 241 EAVNVMPLPLLDRIARIRKTIYVCPMDVIDHDPRFRETQAGDAMRGAFDMEYKRIPIIP 300  
 DB 241 EAVNVMPLPLLDRIARIRKTIYVCPMDVIDHDPRFRETQAGDAMRGAFDMEYKRIPIIP 300  
 QY 301 PELQKADPSDPFESSPVNAGGLFVDRKFWELGQYDGLIEWGBOYEISFKVMCGGRM 360  
 DB 301 PELQKADPSDPFESSPVNAGGLFVDRKFWELGQYDGLIEWGBOYEISFKVMCGGRM 360  
 QY 361 EDIPCSRVGHIYKRYVYKYPAGVSLARNIKRAVAEVMDEYAEITYORREYRLSGDV 420  
 DB 361 EDIPCSRVGHIYKRYVYKYPAGVSLARNIKRAVAEVMDEYAEITYORREYRLSGDV 420  
 QY 421 AVQKLRSSINCKSPKFMFKIAMDLPKFYPPVPPAAGETIRNVTGLCADTKGALG 480  
 DB 421 AVQKLRSSINCKSPKFMFKIAMDLPKFYPPVPPAAGETIRNVTGLCADTKGALG 480  
 QY 481 SPRLBECVARGGEAANNQVFTFWREDIRPDPOHTKKCFDAISHTSPVTLYDCHS 540  
 DB 481 SPRLBECVARGGEAANNQVFTFWREDIRPDPOHTKKCFDAISHTSPVTLYDCHS 540  
 QY 541 MKGNQMLKRYKDKTLVHPVSGSCMDGSESDHRIFMNTCNSSILQOQLFHTNSTVLEKF 600  
 DB 541 MKGNQMLKRYKDKTLVHPVSGSCMDGSESDHRIFMNTCNSSILQOQLFHTNSTVLEKF 600  
 QY 601 NRN 603  
 DB 601 NRN 603

RESULT 2  
 ID\_GLI0\_PAT STANDARD; PRT; 603 AA.  
 AC Q925R7;  
 DT 01-OCT-2004 (Rel. 45, Created)  
 DT 01-OCT-2004 (Rel. 45, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Polypeptide N-acetylglucosaminyltransferase 10 (EC 2.4.1.41)  
 DE (Protein-UDP acetylglucosaminyltransferase 10) (UDP-  
 DE GalNAc:polypeptide N-acetylglucosaminyltransferase 10) (polypeptide  
 DE GalNAc transferase 10) (GalNAc-T10) (p-GANTase 10).  
 GN Name=Galnt10;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., ENZYME ACTIVITY, AND TISSUE SPECIFICITY.  
 RC TISSUE=Sublingual gland;  
 RC MEDLINE=21264503; Pubmed=11278534; DOI=10.1074/jbc.M009638200;  
 RA Ten Hagen K.G., Bedt G.S., Tetaert D., Kingsley P.D., Hagen F.K.,  
 RA Balye M.W., Beres T.M., Degand P., Tabak L.A.;  
 RT "Cloning and characterization of a ninth member of the UDP-  
 RT GalNAc:polypeptide N-acetylglucosaminyltransferase family,"  
 RT pGANTase-79,";  
 RT J. Biol. Chem. 276:17395-17404(2001).  
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has activity toward Muc5Ac and EA2 peptide  
 CC substrates.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium (By similarity).  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
 CC similarity).  
 CC -1- TISSUE SPECIFICITY: Highly expressed in the sublingual gland,  
 CC testis, small intestine, colon and ovary. Expressed at

intermediate level in heart, brain, spleen, lung, stomach, cervix and uterus.

-1- DOMAIN: There are two conserved domains in the glycosyltransferase region: the N-terminal domain (domain A, also called GTI motif), which is probably involved in manganese coordination and substrate binding and the C-terminal domain (domain B, also called Gal/galnac-T motif), which is probably involved in catalytic reaction and UDP-Gal binding (By similarity).

-1- DOMAIN: The ricin B-type lectin domain binds to galnac and contributes to the glycopeptide specificity (By similarity).

-1- SIMILARITY: Belongs to the glycosyltransferase family 2. Galnac-T subfamily.

-1- SIMILARITY: Contains 1 ricin B-type lectin domain.

-1- CAUTION: According Ref.1, this enzyme is unable to transfer Galnac onto serine or threonine residue on the protein receptor, but instead requires the prior addition of a Galnac on a peptide before adding additional Galnac moieties, thereby acting as a glycopeptide transferase.

-1- CAUTION: Was originally (Ref.1) termed Galnac9/ps-GalNase 9.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.iesb-sib.ch/announce/> or send an email to [license@lab-sib.ch](mailto:license@lab-sib.ch)).

-----

EMBL; AF241241; MAK54498.1; -.

DR HSP; P26514; IKNL.

DR RGD; 69409; Galnt10.

DR InterPro; IPR001173; Glyco trans 2.

DR InterPro; IPR008997; RicinB like.

DR InterPro; IPR000772; Ricin B lectin.

DR Pfam; PF00535; Glycosyltransf\_2; 1.

DR Pfam; PF00652; Ricin B lectin; 3.

DR SMART; SM00458; Ricin B lectin; 1.

DR PROSITE; PS50231; RICIN B LECTIN; 1.

KM Calcium; Glycosyltransferase; Golgi stack; Lectin; Manganese;

KW Signal-anchor; Transferase; Transmembrane.

FT DOMAIN 1 11 Cytoplasmic (Potential).

FT TRANSMEM 12 31 Signal-anchor for type II membrane protein (Potential).

FT DOMAIN 32 603 Luminal (Potential).

FT DOMAIN 144 253 Catalytic subdomain A.

FT DOMAIN 311 373 Catalytic subdomain B.

FT DOMAIN 458 590 Ricin B-type lectin.

FT DISULFID 471 488 By similarity.

FT DISULFID 523 538 By similarity.

FT DISULFID 563 578 By similarity.

FT CARBOHYD 124 124 N-linked (GLCNAC. . .) (Potential).

FT CARBOHYD 146 146 N-linked (GLCNAC. . .) (Potential).

FT CARBOHYD 593 593 N-linked (GLCNAC. . .) (Potential).

SEQ SEQUENCE 603 AA; 69116 MW; 194EBA6264ABBF CRC64;

Query March 96.3%; Score 3157; DB 1; Length 603;

Best Local Similarity 96.0%; Pred. No. 2e-247;

Matches 579; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

QY 1 MRRERKLLQAVLALVLLPVGGLMALYREROPDGTGCGSGAANAAPAGQSGHSRQK 60

DB 1 MRRERKLLQAVLALVLLPVGGLMALYREROPDGTGCGSGAANAAPAGQSGHSRQK 60

QY 61 KTFPLGSGQKIKDWHDEAIRDARVNGEGQRPYPTMDAERVDQAAREGFINIYVSDK 120

DB 61 KTFPLGSGQKIKDWHDEAIRDARVNGEGQRPYPTMDAERVDQAAREGFINIYVSDK 120

QY 121 ISLRSRLPDIRHPNCNKRKYLETLPNTSIIIPFNHGGSSLLRTVHSLVNSPELVAEI 180

DB 121 ISLRSRLPDIRHPNCNKRKYLETLPNTSIIIPFNHGGSSLLRTVHSLVNSPELVAEI 180

QY 181 VLVDPSDRHLKKPLEDYNALPSPSVRLIIRTKKRGILRTMLGASVATGVITFLDSHC 240

DB 181 VLVDPSDRHLKKPLEDYNALPSPSVRLIIRTKKRGILRTMLGASVATGVITFLDSHC 240

DB 181 VLVDPSDRHLKKPLEDYNALPSPSVRLIIRTKKRGILRTMLGASVATGVITFLDSHC 240

QY 241 EAVNMLPPLDIRARARKTIVCPMIDVIDHDDPRRYTOAGDAMRGAFDMEYKRIPIR 300

DB 241 EAVNMLPPLDIRARARKTIVCPMIDVIDHDDPRRYTOAGDAMRGAFDMEYKRIPIR 300

QY 301 PELQKADPSDPFESPVNAGGLFAVDKRMFELGQYDGLIWMGEQYISFKVMCGGRM 360

DB 301 PELQKADPSDPFESPVNAGGLFAVDKRMFELGQYDGLIWMGEQYISFKVMCGGRM 360

QY 361 EDIPCSVGHITIKRYIVYKIPAGVSLARNIKRAVAEVMMDYAEITVYRRREYRLSAGDV 420

DB 361 EDIPCSVGHITIKRYIVYKIPAGVSLARNIKRAVAEVMMDYAEITVYRRREYRLSAGDV 420

QY 421 AVQKGLSSINCKSPFKPMFKIAMDLPKPYPPVEPPAAMGEIRNVTGLCAOTKRGALG 480

DB 421 AVQKGLSSINCKSPFKPMFKIAMDLPKPYPPVEPPAAMGEIRNVTGLCAOTKRGALG 480

QY 481 SPLRLFCIRGGEAANNSQVFFFTWRREDIRPDGPQHTKKCFDAVSHSPVLYDCHS 540

DB 481 SPLRLFCIRGGEAANNSQVFFFTWRREDIRPDGPQHTKKCFDAVSHSPVLYDCHS 540

QY 541 MKGNQIMKRYKDKTLVHPVSGSCMDCSBDRIFMNTCNSSSLTQQLFHTNSTVLEKF 600

DB 541 MKGNQIMKRYKDKTLVHPVSGSCMDCSBDRIFMNTCNSSSLTQQLFHTNSTVLENF 600

QY 601 NRN 603

DB 601 NRN 603

RESULT 3

BAD21405 PRELIMINARY; PRT; 634 AA.

ID BAD21405

AC BAD21405; PRELIMINARY; PRT; 634 AA.

DT 01-JUN-2004 (TrEMBLrel. 27, Created)

DT 01-JUN-2004 (TrEMBLrel. 27, Last sequence update)

DT 01-JUN-2004 (TrEMBLrel. 27, Last annotation update)

DE MFLJ00205 protein (fragment).

GN MFLJ00205

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N. A.

RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S., Saga Y., Kitamura H., Nakagawa T., Nagase T., Ohara O., Koga H.; "Prediction of the Coding Sequences of Mouse Homologues of FLJ Gene: The Complete Nucleotide Sequences of 110 Mouse FLJ-homologous cDNAs Identified by Screening of Terminal sequences of cDNA Clones Randomly RT Amplified from Size-Fractionated Libraries."

RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL; AKJ31155; BAD21405.1; -.

FT NON TER 1

SEQ SEQUENCE 634 AA; 72315 MW; C79B82D4D0052C81 CRC64;

Query March 95.7%; Score 3138; DB 2; Length 634;

Best Local Similarity 95.5%; Pred. No. 7.6e-246;

Matches 576; Conservative 9; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRRERKLLQAVLALVLLPVGGLMALYREROPDGTGCGSGAANAAPAGQSGHSRQK 60

DB 32 MRRERKLLQAVLALVLLPVGGLMALYREROPDGTGCGSGAANAAPAGQSGHSRQK 91

QY 61 KTFPLGSGQKIKDWHDEAIRDARVNGEGQRPYPTMDAERVDQAAREGFINIYVSDK 120

DB 92 KTFPLGSGQKIKDWHDEAIRDARVNGEGQRPYPTMDAERVDQAAREGFINIYVSDK 151

QY 121 ISLRSRLPDIRHPNCNKRKYLETLPNTSIIIPFNHGGSSLLRTVHSLVNSPELVAEI 180

DB 152 ISLRSRLPDIRHPNCNKRKYLETLPNTSIIIPFNHGGSSLLRTVHSLVNSPELVAEI 211



QY 181 VLVDSDREHLLKKPLEDYALFPSPVRLTKRKGILRTKMLGASVATGDTVFLDSHC 240  
 Db 212 VLVDSDREHLLKKPLEDYALFPSPVRLTKRKGILRTKMLGASVATGDTVFLDSHC 271  
 QY 241 EAVNNVLPILLDIRARRKRTIVCEMIDVIDHDDFRYETQAGDAMRGAEDWEMYYKRIPIP 300  
 Db 272 EAVNNVLPILLDIRARRKRTIVCEMIDVIDHDDFRYETQAGDAMRGAEDWEMYYKRIPIP 331  
 QY 301 PELQKADSPDPESPVPWAGGLFANDRKFWELGSDGLTWGGEQVEISFKVMCCGRM 360  
 Db 332 PELQKADSPDPESPVPWAGGLFANDRKFWELGSDGLTWGGEQVEISFKVMCCGRM 391  
 QY 361 EDIPCSRVGHTYRKYVYKYPAGVSLARNLKRVAVNMDEYAEITYORRPEYRLSAGDV 420  
 Db 392 EDIPCSRVGHTYRKYVYKYPAGVSLARNLKRVAVNMDEYAEITYORRPEYRLSAGDV 451  
 QY 421 AVQKRLSSLNCKSPKFMFKTIANDLKPYPPVPPAAANGELRNVTGLCADTKHAGLG 480  
 Db 452 VAQKRLRVSLNCKSPKFMFKTIANDLKPYPPVPPAAANGELRNVTGLCADTKHAGLG 511  
 QY 481 SPRLBECVGRGHAANNNOVFTFTWRREDIRPEDPQHTKFCPCDAISHTSPVTLNCHS 540  
 Db 512 SPRLBECVGRGHAANNNOVFTFTWRREDIRPEDPQHTKFCPCDAISHTSPVTLNCHS 571  
 QY 541 MKGNOLMKYRKDKTLVHPVSGSCMDSPDRIFPMTCPNSLSLQOQLFHTNSTVLEKF 600  
 Db 572 MKGNOLMKYRKDKTLVHPVSGSCMDSPDRIFPMTCPNSLSLQOQLFHTNSTVLEKF 631  
 QY 601 NRN 603  
 Db 632 NRN 634

# RESULT 4 GL10 MOUSE STANDARD; PRF; 603 AA.

ID GL10 MOUSE STANDARD; PRF; 603 AA.  
 AC 06P957; 06KAO2; 08BZ08; 091YD6;  
 DT 01-OCT-2004 (Rel. 45, Created)  
 DT 01-OCT-2004 (Rel. 45, Last sequence update)  
 DE Polypeptide N-acetylgalactosaminyltransferase 10 (EC 2.4.1.41)  
 DE (Protein-UDP-acetylgalactosaminyltransferase 10) (UDP-  
 GAlNAc:polypeptide N-acetylgalactosaminyltransferase 10) (polypeptide  
 GAlNAc transferase 10) (GAlNAc-T10) (pp-GalTase 10).  
 GN Name=GAlnT10;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,  
 RA Suga Y., Kitamura H., Nakagawa T., Nagase T., Ohara O., Koga H.;  
 RA "Prediction of the coding sequences of mouse homologues of Flj genes:  
 RT the complete nucleotide sequences of 110 mouse Flj-homologous cDNAs  
 RT identified by screening of terminal sequences of cDNA clones randomly  
 RT sampled from size-fractionated libraries.";  
 RL DNA Res. 11:167-180(2004).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6, and FVB/N. TISSUE=Brain, and Breast tumor;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Collins B.A., Grouse L.H., Derge J.G.,  
 RA Klannberg R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh P.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Marisano K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stropstein M., Soares M.B., Bonaldo M.F., Casavant T.J., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loughran N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Wooley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,  
 RA Butterfield Y.S.N., Krzywicki M.I., Skalka U., Smalins D.E.,  
 RA Scherch A., Schein J.B., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 76-603 FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Colon;  
 RX MEDLINE=22346683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaido I., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schombach C., Gotoori T.,  
 RA Baldeceli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schirral L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Choitha C., Corbani L.E., Cousins S.,  
 RA Datta E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okito T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.V., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada M.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verrardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wymshaw-Boris A., Yanagisawa M., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume M., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yamanishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP TISSUE SPECIFICITY  
 RX MEDLINE=21264503; PubMed=11278534; DOI=10.1074/jbc.M009638200;  
 RA Ten Hagen K.G., Bedi G.S., Tetaert D., Kingsley P.D., Hagen F.K.,  
 RA Balys M.M., Beres T.M., Degand P., Tabak L.A.;  
 RT "Cloning and characterization of a ninth member of the UDP-  
 RT GAlNAc:polypeptide N-acetylgalactosaminyltransferase family,  
 RT ppGalTase-19.";  
 RL J. Biol. Chem. 276:17395-17404(2001).  
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has activity toward Muc8Ac and E2 peptide  
 CC substrates (By similarity).  
 CC CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium (By similarity).  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
 CC similarity).  
 CC -1- TISSUE SPECIFICITY: Expressed at higher level than GAlNT9. In the  
 CC developing hindbrain region of E14.5 embryos it accumulates in the  
 CC rapidly dividing, undifferentiated ventricular zone adjacent to  
 CC the pons. It also accumulates in the regions immediately rostral  
 CC and caudal to the dorsal rhombic lips differentiating into the  
 CC cerebellum. Not expressed in the developing choroid plexus.  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called G1 motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC Gal/GalNAc-T motif), which is probably involved in catalytic

```

CC reaction and UDP-Gal binding (By similarity).
CC - DOMAIN: The ricin B-type lectin domain binds to GAlNAc and
CC contributes to the glycopeptide specificity (By similarity).
CC - SIMILARITY: Belongs to the glycosyltransferase family 2. GAlNAc-T
CC subfamily.
CC - SIMILARITY: Contains 1 ricin B-type lectin domain.
CC CAUTION: According to experiments made in rat, this enzyme is
CC unable to transfer GAlNAc onto serine or threonine residue on the
CC protein receptor, but instead requires the prior addition of a
CC GAlNAc on a peptide before adding additional GAlNAc moieties,
CC thereby acting as a glycopeptide transferase.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1db-sib.ch/announce/
CC or send an email to license@1db-sib.ch).
CC -----
DR EMBL: AK131155; BAD21405.1; ALT_INIT.
DR EMBL: BC016585; AAH16585.1; ALT_INIT.
DR EMBL: BC060617; AAH60617.1; -.
DR EMBL: AK033515; BAC28334.1; -.
DR HSBP: P26514; 1KIM.
DR MGD: MGI:1890480; Galnt10.
DR GO: GO:0004653; F:polyepide N-acetylgalactosaminyltransferase. . .; IDA.
DR GO: GO:0006493; P:O-linked glycosylation; IDA.
DR InterPro: IPR001173; Glyco_trans_2.
DR InterPro: IPR008997; RicinB_like.
DR Pfam: PF00535; Glycos_transf_2; 1.
DR Pfam: PF00652; Ricin B lectin; 2.
DR SMART: SM00458; RICIN_1.
DR PROSITE: PS00231; RICIN_B_LECTIN; 1.
DR Calxim; Glycosyltransferase; Golgi stack; lectin; Manganese;
KW Signal-anchor; transferase; Transmembrane.
FT DOMAIN 1 11 Cytoplasmic (Potential).
FT TRANSMEM 12 31 Signal-anchor for type II membrane
FT 32 603 protein (Potential).
FT DOMAIN 144 253 Luminal (Potential).
FT DOMAIN 311 373 Catalytic subdomain A.
FT DOMAIN 458 590 Catalytic subdomain B.
FT DISULFID 471 488 Ricin B-type lectin.
FT DISULFID 523 538 By similarity.
FT DISULFID 563 578 By similarity.
FT CARBOHYD 124 124 N-linked (GlcNAc . .) (Potential).
FT CARBOHYD 146 146 N-linked (GlcNAc . .) (Potential).
FT CARBOHYD 593 593 N-linked (GlcNAc . .) (Potential).
FT CONFLICT 233 233 V -> I (in Ref. 1).
SQ SEQUENCE 603 AA; 69116 MW; PF55FBA7E1DD7544 CRC64;

Query Match 95.7%; Score 3137; DB 1; Length 603;
Best Local Similarity 95.4%; Pred. No. 8.6e-246;
Matches 575; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

```

```

QY 241 EAVNVMPLPDLRIARRKTIICVEMIDVIDHDDFRFETQAGADAMRGAFDMEMYKRIPIIP 300
DB 241 EAVNVMPLPDLRIARRKTIICVEMIDVIDHDDFRFETQAGADAMRGAFDMEMYKRIPIIP 300
QY 301 PELQKADPSPPSPSPVWAGGLFAVDRKFWELGSDGLTIWGEQVEISFKVMCCGRM 360
DB 301 PELQKADPSPPSPSPVWAGGLFAVDRKFWELGSDGLTIWGEQVEISFKVMCCGRM 360
QY 361 EDIPCSRVGHYIRYCVYKYYPAGVSLARNLKRVAEVWMDVYAEYIYORREYRLISAGDV 420
DB 361 EDIPCSRVGHYIRYCVYKYYPAGVSLARNLKRVAEVWMDVYAEYIYORREYRLISAGDV 420
QY 421 AVQKGLRSSLNCKSEFKFMFKTIAMDLPKYPVPEPPAAANGELRNVTGLCADTKRGALG 480
DB 421 AVQKGLRSSLNCKSEFKFMFKTIAMDLPKYPVPEPPAAANGELRNVTGLCADTKRGALG 480
QY 481 SPLRLKECVGRGSAANMNOVFTFRREDIRPDPQHTKKFCDAISHSPTLYCHS 540
DB 481 SPLRLKECVGRGSAANMNOVFTFRREDIRPDPQHTKKFCDAISHSPTLYCHS 540
QY 541 MKGNQMKYRKDKTLVHPVSGSCMDSCSDHRIIPMTCNPSLSLTOQLFEHTNSTVLEKF 600
DB 541 MKGNQMKYRKDKTLVHPVSGSCMDSCSDHRIIPMTCNPSLSLTOQLFEHTNSTVLEKF 600
QY 601 NRN 603
DB 601 NRN 603

RESULT 5
AAH60617 PRELIMINARY; PRT; 603 AA.
AC AAH60617;
DT 12-MAY-2004 (TREMblrel. 27, Created)
DT 12-MAY-2004 (TREMblrel. 27, Last sequence update)
DT 12-MAY-2004 (TREMblrel. 27, Last annotation update)
DE UDP-N-acetyl-alpha-D-galactosamine:polypeptide
DE N-acetylgalactosaminyltransferase 10.
GN GALNT10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22386257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Boeak S.A., McGwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Boulford G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Touchman J.W., Schmitt J., Myers R.M., Butterfield Y.S.,
RA Krywinski M.I., Skelton U., Smalton D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strauberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC060617; AAH60617.1; -.

```



KM Transferrase.  
 SQ SEQUENCE 603 AA; 69116 MW; PF55FBA7B1DD7544 CRC64;  
 Query Match 95.7%; Score 3137; DB 2; Length 603;  
 Best Local Similarity 95.4%; Pred. No. 8.6e-246;  
 Matches 575; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRREKRLDVALVLAALVLPVNGMALYREORPOTGGSGAAAPAAAGCSHRÖK 60  
 DB 1 MRREKRLDVALVLAALVLPVNGMALYREORPOTGGSGAAAPAAAGCSHRÖK 60  
 QY 61 KTFEFLGQGLKQWMDKEAIRDAQVNGEGGRPYPM-TDAERVDQYRENGFYISDK 120  
 DB 61 KTFEFLGQGLKQWMDKEAIRDAQVNGEGGRPYPM-TDAERVDQYRENGFYISDK 120  
 QY 121 ISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEGSSILRTVHSLNRSPELVAEI 180  
 DB 121 ISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEGSSILRTVHSLNRSPELVAEI 180  
 QY 181 VLVDPSDRHLKKPLLEDYMA-LFPSVRIARTKRREGLITRMLGASVATGDTFLDSHC 240  
 DB 181 VLVDPSDRHLKKPLLEDYMA-LFPSVRIARTKRREGLITRMLGASVATGDTFLDSHC 240  
 QY 241 EAVNMLPPLDIRIARAKTIVCPMIDVIDHDPRYETOAGDAMRGAFPMEMYKRIPIP 300  
 DB 241 EAVNMLPPLDIRIARAKTIVCPMIDVIDHDPRYETOAGDAMRGAFPMEMYKRIPIP 300  
 QY 301 PELÖKADSPDPFESPVNAGGLFAVDRKFWELGSDPGLIWGEGÖYEISFKVMCGGRM 360  
 DB 301 PELÖKADSPDPFESPVNAGGLFAVDRKFWELGSDPGLIWGEGÖYEISFKVMCGGRM 360  
 QY 361 EDIPCSVGHITRYKYVPYKVPAGVSLARNLKRVAEVMMDYAEITTYORREYRLSGDV 420  
 DB 361 EDIPCSVGHITRYKYVPYKVPAGVSLARNLKRVAEVMMDYAEITTYORREYRLSGDV 420  
 QY 421 AVOKLRSLSNCSFKFMFTKIAMDLPKYPPVPEPAAWGEINNGVGLCADRKHALG 480  
 DB 421 AVOKLRSLSNCSFKFMFTKIAMDLPKYPPVPEPAAWGEINNGVGLCADRKHALG 480  
 QY 481 SPLRLGCGVGRGEAAMNNQVFTFTWRBDIRPGDPQHTKKCFDASISHTSPVTLYDCHS 540  
 DB 481 SPLRLGCGVGRGEAAMNNQVFTFTWRBDIRPGDPQHTKKCFDASISHTSPVTLYDCHS 540  
 QY 541 MKGNOLMKYRKDKTLTHPVSGSCMDCSBDRHIFMNTCNPSLTQOQLFHTNSTVLEKF 600  
 DB 541 MKGNOLMKYRKDKTLTHPVSGSCMDCSBDRHIFMNTCNPSLTQOQLFHTNSTVLEKF 600  
 QY 601 NRN 603  
 DB 601 NRN 603  
 QY 601 NRN 603  
 DB 601 NRN 603

RESULT 6  
 Q700E9 PRELIMINARY; PRT; 644 AA.  
 AC Q700E9; PRELIMINARY; PRT; 644 AA.  
 DT 01-MAR-2004 (TEMBLrel. 26, Created)  
 DT 01-MAR-2004 (TEMBLrel. 26, last sequence update)  
 DT 01-MAR-2004 (TEMBLrel. 26, last annotation update)  
 DE AGCP9480 (Fragment).  
 GN Name=agc9480; ORFNames=ENSG00000009226;  
 OS Anopheles gambiae str. PEST.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
 NCBI\_TaxID=180454;  
 RN NCBI\_TaxID=180454;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.

DR EMBL; AA001008986; EAA00190.1; -;  
 DR InterPro; IPR001172; Glyco\_transf. 2.  
 DR InterPro; IPR007772; Ricin\_B\_lectin.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR Pfam; PF00652; Ricin\_B\_lectin; 3.  
 DR PROSITE; PS50231; RICIN\_B\_LECTIN; 1.  
 FT NON TER 1  
 SQ SEQUENCE 644 AA; 74214 MW; 0B7AB007BC228501 CRC64;  
 Query Match 42.8%; Score 1404; DB 2; Length 644;  
 Best Local Similarity 45.6%; Pred. No. 5.6e-105;  
 Matches 283; Conservative 104; Mismatches 195; Indels 38; Gaps 16;

QY 1 MRREKRLDVALVLAALVLPVNGMALYREORPOTGGSGAAAPAAAGCSHRÖK 60  
 DB 1 MRREKRLDVALVLAALVLPVNGMALYREORPOTGGSGAAAPAAAGCSHRÖK 60  
 QY 61 KTFEFLGQGLKQWMDKEAIRDAQVNGEGGRPYPM-TDAERVDQYRENGFYISDK 113  
 DB 61 KTFEFLGQGLKQWMDKEAIRDAQVNGEGGRPYPM-TDAERVDQYRENGFYISDK 113  
 QY 90 EGSEFNNPSKAVHOKRIDMANYDLIHEAKRSGIGEGKAGQDKSEHMKDKLFXKNGF 149  
 DB 90 EGSEFNNPSKAVHOKRIDMANYDLIHEAKRSGIGEGKAGQDKSEHMKDKLFXKNGF 149  
 QY 114 NIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEGSSILRTVHSLNRS 173  
 DB 114 NIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEGSSILRTVHSLNRS 173  
 QY 150 NAVLSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEGSSILRTVHSLNRS 209  
 DB 150 NAVLSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEGSSILRTVHSLNRS 209  
 QY 174 PELVAEIVLVDPSDRHLKKPLLEDYMA-LFPSVRIARTKRREGLITRMLGASVATGDT 232  
 DB 174 PELVAEIVLVDPSDRHLKKPLLEDYMA-LFPSVRIARTKRREGLITRMLGASVATGDT 232  
 QY 210 PELVAEIVLVDPSDRHLKKPLLEDYMA-LFPSVRIARTKRREGLITRMLGASVATGDT 269  
 DB 210 PELVAEIVLVDPSDRHLKKPLLEDYMA-LFPSVRIARTKRREGLITRMLGASVATGDT 269  
 QY 233 ITFLDSHCANVMMLPPLDIRIARAKTIVCPMIDVIDHDPRYETOAGDAMRGAFPMEM 292  
 DB 233 ITFLDSHCANVMMLPPLDIRIARAKTIVCPMIDVIDHDPRYETOAGDAMRGAFPMEM 292  
 QY 270 LIFLDSHCANVMMLPPLDIRIARAKTIVCPMIDVIDHDPRYETOAGDAMRGAFPMEM 328  
 DB 270 LIFLDSHCANVMMLPPLDIRIARAKTIVCPMIDVIDHDPRYETOAGDAMRGAFPMEM 328  
 QY 293 YKRRIP-IPPELOKADSPDPFESPVNAGGLFAVDRKFWELGSDPGLIWGEGÖYEISF 351  
 DB 293 YKRRIP-IPPELOKADSPDPFESPVNAGGLFAVDRKFWELGSDPGLIWGEGÖYEISF 351  
 QY 329 FYRRLPLPDLQ-NTTEFESPVNAGGLFAVDRKFWELGSDPGLIWGEGÖYEISF 386  
 DB 329 FYRRLPLPDLQ-NTTEFESPVNAGGLFAVDRKFWELGSDPGLIWGEGÖYEISF 386  
 QY 352 KVMCGGRMDIPCSRGHITRYKYVPYKVPAGVSLARNLKRVAEVMMDYAEITTYOR-R 409  
 DB 352 KVMCGGRMDIPCSRGHITRYKYVPYKVPAGVSLARNLKRVAEVMMDYAEITTYOR-R 409  
 QY 387 KIMCGGKMDADCSRGHITRYKYVPYKVPAGVSLARNLKRVAEVMMDYAEITTYOR-R 446  
 DB 387 KIMCGGKMDADCSRGHITRYKYVPYKVPAGVSLARNLKRVAEVMMDYAEITTYOR-R 446  
 QY 410 PEYRLSAGVAVOKLRSLSNCSFKFMFTKIAMDLPKYPPVPEPAAWGEINNGVGL 468  
 DB 410 PEYRLSAGVAVOKLRSLSNCSFKFMFTKIAMDLPKYPPVPEPAAWGEINNGVGL 468  
 QY 447 KTYENDVGDISRQLAIREKLOCFKFMFTKIAMDLPKYPPVPEPAAWGEINNGVGL 506  
 DB 447 KTYENDVGDISRQLAIREKLOCFKFMFTKIAMDLPKYPPVPEPAAWGEINNGVGL 506  
 QY 469 GLCADTGGALGSPRLGCGVGRGEAAMNNQVFTFTWRBDIRPGDPQHTKKCFDASISHTSP 525  
 DB 469 GLCADTGGALGSPRLGCGVGRGEAAMNNQVFTFTWRBDIRPGDPQHTKKCFDASISHTSP 525  
 QY 507 ALCVDTLNHGEKQTLIGYSCADKQOPN-QFOLSMHRLR-----IKGELCMD 557  
 DB 507 ALCVDTLNHGEKQTLIGYSCADKQOPN-QFOLSMHRLR-----IKGELCMD 557  
 QY 526 AISHTSP---VTLYDCHSMGNOLMKYRKDKTLTHPVSGSCMDCSBDRHIFMNTCNPS 581  
 DB 526 AISHTSP---VTLYDCHSMGNOLMKYRKDKTLTHPVSGSCMDCSBDRHIFMNTCNPS 581  
 QY 558 -VSESVNAXKILLYHCHGCGQGNQIMRYEPTQMLKQGNRCLDMNBNNEVFVNPCBPT 616  
 DB 558 -VSESVNAXKILLYHCHGCGQGNQIMRYEPTQMLKQGNRCLDMNBNNEVFVNPCBPT 616  
 QY 582 SLTQOQLFHTNSTVLEKN 601  
 DB 582 SLTQOQLFHTNSTVLEKN 601  
 QY 617 NPROKMRWGFINATSLQMN 636  
 DB 617 NPROKMRWGFINATSLQMN 636

RESULT 7  
 GL10 CAEBL STANDARD; PRT; 622 AA.  
 AC 045947;  
 DT 01-OCT-2004 (Rel. 45, Created)  
 DT 01-OCT-2004 (Rel. 45, last sequence update)  
 DT 01-OCT-2004 (Rel. 45, last annotation update)  
 DE Putative polypeptide N-acetylglucosaminyltransferase 10  
 DE (EC 2.4.1.41) (Protein-UDP-acetylglucosaminyltransferase 10) (UDP-  
 DE GalNAc:polypeptide N-acetylglucosaminyltransferase 10) (pp-GaNTase  
 DE 10).  
 GN Name=gl10; ORFNames=Y45F10D.3;  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;  
 OC Rhabditidae; Peleodierinae; Caenorhabditis.

OX NCBI\_TaxID=6239;  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA McMurtry A.A.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: May catalyze the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor (By similarity).  
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- Cofactor: Manganese and calcium (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
 CC similarity).  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called GT1 motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC Gal/GalNAc-T motif), which is probably involved in catalytic  
 CC reaction and UDP-Gal binding (By similarity).  
 CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and  
 CC contributes to the glycopeptide specificity (By similarity).  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
 CC subfamily.  
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage fee and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL: AL021492; CAA16378.1; AL1\_INIT.  
 DR PIR: T26930; T26930.  
 DR Wormpep: Y45F10D.3; CB16642.  
 DR InterPro: IPR001173; Glyco\_trans\_2.  
 DR InterPro: IPR008997; RicinB-like.  
 DR InterPro: IPR000772; RicinB\_lectin.  
 DR Pfam: PF00535; Glycos\_transf\_2; 1.  
 DR Pfam: PF00652; RicinB\_lectin; 3.  
 DR SMART: SMO0458; RICIN\_1.  
 DR PROSITE: PS50231; RICIN\_B\_LECTIN; 1.  
 KM Calcium; Glycosyltransferase; Golgi stack; Hypothetical protein;  
 FT Lectin; Manganese; Signal-anchor; Transferase; Transmembrane.  
 FT DOMAIN 1 12  
 FT TRANSMEM 13 32  
 FT SIGNAL-ANCHOR FOR TYPE II MEMBRANE  
 FT PROTEIN (Potential).  
 FT LUMENAL (Potential).  
 FT DOMAIN 156 622  
 FT DOMAIN 324 386  
 FT DOMAIN 519 622  
 FT DISULFID 532 549  
 FT DISULFID 575 591  
 FT CARBOHYD 136 136  
 FT CARBOHYD 170 170  
 FT N-linked (GlcNAc...) (Potential).  
 FT N-linked (GlcNAc...) (Potential).  
 SQ SEQUENCE 622 AA; 71811 MW; FBA3A21159C7E452 CRC64;  
 Query Match 42.5%; Score 1392.5; DB 1; Length 622;  
 Best Local Similarity 48.8%; Pred. No. 4.5e-104;  
 Matches 279; Conservative 87; Mismatches 175; Indels 31; Gaps 14;  
 OY 49 PAAGSGSHRQKTFLLDQGLKDMHDKAIRDPAQVNGEGQRPMTDAERVD-- 106  
 DB 59 PPAALGDALDLPFEKRGH-EKIK-WEDEAAYEKREKREGGEMKPYKLPEDKEVERKA 116  
 OY 107 --ATRENGFNIVGDKISLNSLPRIKPNCKSKRYLTLPTPTSLIIFPNHGMSSLLRT 164  
 DB 117 LSLVANGYNAVYISDMISLNSIKDIRKECKNNMYSAKLPTVSIVLPFHEHNSHTLRS 176

OY 165 VHSVYLNRPPELVVAELVLDPSDRHLKKPLEDYM---ALPESVRLIRTKREGILRTR 221  
 DB 177 VYSVYNSPPELKEILITVDPSSEKPALRQPLEDFLKKNKIDHIVKLTREKREGILRGR 236  
 OY 222 MLAGSVATGVDITFLDHCANVMMLPPLDRIARNKRTVCPMIDVIDHDDPFYERQAG 281  
 DB 237 QLGQADVTGELILFLDHSBANVMWLPPLDPAEDRYRTVCPVDVIDCTEYRPO-D 295  
 OY 282 DAMGAFDEMYVYKRIPIPELQKADSDPFESVVMAGLFAVDKRMFWEIGYDPGLBI 341  
 DB 296 BGARSGFDMAFNRYRLTLTK-DRESPTKFNPSVVMAGFALSAKFWELGCGYDEGLDI 354  
 OY 342 WGEQVEYISFKVMCGRMEDIPCSRVGHIYR-KTVYK-VPGVSLARLKVAAEYVMD 399  
 DB 355 WGEQVEYISFKVMCGRMEDIPCSRVGHIYR-KTVYK-VPGVSLARLKVAAEYVMD 414  
 OY 400 EYAEYIYQRPEPYHLSAGVAVQKLRSSLNCSFPMFTKTLAMDLPKFPYPPPPAAA 459  
 DB 415 DYKETLYKRRGVNNAADGLKMKGRKLOCKSPFMKEIAVDDKTYPAVEPKASA 474  
 OY 460 WGEIRNYGTGLCADTKGALGSPLEEGVYR---GRGEAAMNNMVFYFTMRDIPGD 515  
 DB 475 EGEIRNYGTGLCADTKGALGSPLEEGVYR---GRGEAAMNNMVFYFTMRDIPGD 526  
 OY 516 PQHTKKECFPAISHT--SPVTLVYDCHSMKGNQMKYR-KQTLVHPVSGSCMDCSSDHR 572  
 DB 527 --KRRKICFDGSTVDKAPVILFPCHSKMGQLEFKYVAQIQIYHPISSGCLTADENKGG 584  
 OY 573 -IFNATCNPSGLTQOMLFHTNSTVLEKFNEN 603  
 DB 585 FLHMKCDSSDQLQKMWAMQVNLLETRQAN 616

## RESULT 8

CLT6\_DROME STANDARD; PRT; 666 AA.  
 ID CLT6\_DROME  
 AC Q6WV16; Q95R40; Q9VZX5;  
 DT 01-OCT-2004 (Ref. 45, Created)  
 DT 01-OCT-2004 (Ref. 45, Last sequence update)  
 DT 01-OCT-2004 (Ref. 45, Last annotation update)  
 DE N-acetylgalactosaminyltransferase 6 (EC 2.4.1.-) (Protein-UDP  
 DE acetylgalactosaminyltransferase 6) (UDP-GalNAc:polypeptide N-  
 DE acetylgalactosaminyltransferase 6) (pp-GalNAc 6).  
 GN Name=pgant6; ORFNames=CG2103;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A., ENZYME ACTIVITY, TISSUE SPECIFICITY, AND  
 RP DEVELOPMENTAL STAGE.  
 RC STRAIN=Canton-S; TISSUE=Embryo;  
 RX MEDLINE=22841110; PubMed=12829714; DOI=10.1074/jbc.M303836200;  
 RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.;  
 RA "Functional characterization and expression analysis of members of the  
 RA UDP-GalNAc:polypeptide N-acetylglactosaminyltransferase family from  
 RA Drosophila melanogaster.";  
 RL J. Biol. Chem. 278:35039-35048 (2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=Berkeley;  
 RC MEDLINE=20196006; PubMed=10731133; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoekstra R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.,  
 RA Abtill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolintsov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Butrie K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
RA de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fodor C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,  
RA Jostali M., Kalush F., Karen G.H., Ke Z., Kemion J.A., Kechum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulj D., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Jiang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobery C., Morris C., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,  
RA Svirskas R., Tector C., Turner R., Venter E.H., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao O.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,  
RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [3]  
RP SEQUENCE FROM N. A.  
RP STRAIN=Berkley; TISSUE=Embryo;  
RC MEDLINE=22426066; PubMed=12537559;  
RX Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
RA George R.A., Guarin H., Krommiller B., Pacle J.M., Park K.H.,  
RA Rubin G.M., Celniker S.E.;  
RT "A *Drosophila* full-length cDNA resource.";  
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
CC -1- FUNCTION: Glycopeptide transferase involved in O-linked  
CC oligosaccharide biosynthesis, which catalyzes the transfer of an  
CC N-acetyl-D-galactosamine residue to an already glycosylated  
CC peptide. In contrast to other proteins of the family, it does not  
CC act as a peptide transferase that transfers GalNAc onto serine or  
CC threonine residue on the protein receptor, but instead requires  
CC the prior addition of a GalNAc on a peptide before adding  
CC additional GalNAc moieties. Some peptide transferase activity is  
CC however not excluded, considering that its appropriate peptide  
CC substrate may remain unidentified. Prefers the diglycosylated  
CC Muc5AC-3/13 as substrate.  
CC -1- COFACTOR: Manganese and calcium (by similarity).  
CC -1- PATHWAY: Glycosylation.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (by  
CC similarity).  
CC -1- TISSUE SPECIFICITY: In embryos, it is specifically expressed in  
CC the salivary glands from stage 12, becoming stronger at stage 13.  
CC Not expressed in other tissues. Further expressed during  
CC oogenesis, in the somatically derived follicle cells that surround  
CC the developing oocyte, which are involved in the maturation of the  
CC oocyte and construction of the egg shell, as well as playing a  
CC role in subsequent embryonic pattern formation.  
CC -1- DEVELOPMENTAL STAGE: Expressed throughout embryonic, larval, pupal  
CC and adult stages, with increasing levels during larval  
CC development.  
CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
CC region: the N-terminal domain (domain A, also called GT1 motif),  
CC which is probably involved in manganese coordination and substrate  
CC binding and the C-terminal domain (domain B, also called  
CC Gal/GalNAc-T motif), which is probably involved in catalytic  
CC reaction and UDP-Gal binding (by similarity).  
CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and  
CC contributes to the glycopeptide specificity (by similarity).  
CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
CC subfamily.  
CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AY268067; AAC56703.1; -
DR EMBL; AE003476; AAP47690.1; -
DR EMBL; AY061629; AAL29177.1; -
DR Flybase; FBgn005375; pgn6.
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR008997; RicinB_1ke.
DR InterPro; IPR00772; RicinB_lectin.
DR Pfam; PF00535; Glycos_transf_2; 1.
DR Pfam; PF00652; RicinB_lectin; 3.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS0231; RICIN_B_LECTIN; 1.
KW Calcium; Glycosyltransferase; Golgi stack; lectin; Manganese;
KW Signal-anchor; Transferase; Transmembrane.
FT DOMAIN 1 11 Cytoplasmic (potential).
FT TRANSMEM 12 31 Signal-anchor for type II membrane
FT FT protein (potential).
FT DOMAIN 201 311 Luminal (potential).
FT FT 367 429 Catalytic subdomain A.
FT DOMAIN 518 648 Catalytic subdomain B.
FT FT 531 548 Ricin B-type lectin.
FT DISULFID 577 594 By similarity.
FT FT 621 636 By similarity.
FT DISULFID 181 181 N-linked (GlcNAc . .) (Potential).
FT CARBOHYD 285 285 N-linked (GlcNAc . .) (Potential).
FT CARBOHYD 651 651 N-linked (GlcNAc . .) (Potential).
FT CARBOHYD 657 657 N-linked (GlcNAc . .) (Potential).
FT CONFLICT 95 96 EV -> DA (in Ref. 1).
FT CONFLICT 107 107 Q -> R (in Ref. 1 and 3).
FT CONFLICT 499 499 E -> K (in Ref. 3).
SQ SEQUENCE 666 AA; 76972 MW; CAECA6CE84860600C CRC64;

Query Match 41.5%; Score 1360; DB 1; Length 666;
Best Local Similarity 49.0%; Pred. No. 2.2e-101;
Matches 277; Conservative 76; Mismatches 160; Indels 52; Gaps 15;

67 DGQKLKMHDKEARLRDAPQRYGNEQGRPYPMTDABRVDAQR---ENGFPIYVSDKSL 123
| | | | | : : : | | | | | : : : | | | | | : : : | | |
121 DASVYKMDHDTTFEKKDKQVGLSGGASTLDDSGRDLEKRWSLSEGFALLSDSISV 180
| | | | | : : : | | | | | : : : | | | | | : : : | | |
124 NRSLLPDIRHPNCSKRYLETLPNLSIIIPHNQMSLLPTVHSVLSNRSPELVAEIVLV 183
| | | | | : : : | | | | | : : : | | | | | : : : | | |
181 NRSVPDIRHPICRKKQEVYAKLPYTSVIIIFNVELSVLSVMSVSLNRSPELMKEIILV 240
| | | | | : : : | | | | | : : : | | | | | : : : | | |
184 DDFSDEHLKPLDEDYNA-LFSPYRIILFYTKRBSGLITRMGLASVATGDVITPLDSCEA 242
| | | | | : : : | | | | | : : : | | | | | : : : | | |
241 DDHSDREYLGKELLETYYIAEHFKVMYRVRLPRRTGLIGARAAGANAAVAEVLIFDLSHVEA 300
| | | | | : : : | | | | | : : : | | | | | : : : | | |
243 NVNMLPRLDIDIAANRKTIVCPMTDIVDDHDFRERYTQAGDMRGAPEWMEYKRIPIPE 302
| | | | | : : : | | | | | : : : | | | | | : : : | | |
301 NVNMLPRLLEIATLNKRTAVCPPLDIVDHTNFHRAQ-DEAKARAPMEEFYKRLPLPE 359
| | | | | : : : | | | | | : : : | | | | | : : : | | |
303 LQKADSPDFSPVMAAGGLFVADRKFMWELGCGYDGLIEIWGSEQEYISFKVMWCGHMEQ 362
| | | | | : : : | | | | | : : : | | | | | : : : | | |
360 DLK-LHPADPFPSPIMAGGLVPIISSEFFWELGCGYDGLDINGBQYELSPKIMWCGEMTD 418
| | | | | : : : | | | | | : : : | | | | | : : : | | |
363 IPCSRVGHYIR-----KYVPYKDVAGVSLANRLKRVAEVWMEYAEYIYQRRPE-YHLS 416
| | | | | : : : | | | | | : : : | | | | | : : : | | |
419 APCSRIGIHYIGRPNHQPSPRK---GDYLNKYKRVAEVWMEDEVKNTLYSHGDLYESVD 475
| | | | | : : : | | | | | : : : | | | | | : : : | | |
417 AGDAVAVQKGLRSSLNCSEFKKFMFKIMMDLPKFPVPEPAAAMGELIRNYGT-GLCADT- 474
| | | | | : : : | | | | | : : : | | | | | : : : | | |
476 PGDLTEQKAITRTKNCSEFKFMEEVAFDLMKITYPPVDPSPSYAAGLQNVGNOMLCEDTL 535
| | | | | : : : | | | | | : : : | | | | | : : : | | |

```

Qy 475 ---RUGALG-----SPLRLEGCVARGEAAANNNOVFTFWREDIRPDPQHTKKF 522  
 Db 536 GRKGNKMGYACADNITKFOR-----TQFWELSKRDLR---LRRKKE 576  
 Qy 523 CFDA--ISHSPVLYDCHSKGNQLMKY-RKDITLYHPVG-SCMCSSESDHRIFMNTC 578  
 Db 577 CLDVQIWDANAPVWLMCHSGGQNGQYVYDRKHQKHGTGRCCLLELPPSGEVANNC 636  
 Qy 579 NPSSLTQOQWLFPHNTSSTVLEKENRN 603  
 Db 637 DTDNRFOQWNGSFNKTALDNYSD 661

## RESULT 9

AAQ56703 PRELIMINARY; PRT; 666 AA.  
 ID AAQ56703  
 AC AAQ56703  
 DT 02-MAR-2004 (TREMBlrel. 27, Created)  
 DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)  
 DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)  
 DE UDP-GalNAc:polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41).  
 GN PGANT6.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=12829714;  
 RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.;  
 RT "Functional Characterization and Expression Analysis of Members of the  
 RT UDP-GalNAc:Polypeptide N-Acetylglactosaminyltransferase Family from  
 RT Drosophila melanogaster";  
 RU J. Biol. Chem. 278:35039-35048(2003).  
 DR EMBL; AY268067; AAQ56703.1; -.  
 KW Glycosyltransferase; Transferase.  
 SQ SEQUENCE 666 AA; 76958 MW; 1D8362E8D8D196 CRC64;

Query Match 41.5%; Score 1360; DB 2; Length 666;  
 Best Local Similarity 49.0%; Pred. No. 2.2e-101; Indels 52; Gaps 15;  
 Matches 277; Conservative 76; Mismatches 160;  
 Qy 67 DGQKLDKMDKREAIIRDAQVNGEGRRPMTDAERVDQVR---ENGFNIVYSDKISTL 123  
 Db 121 DASVKKMDHDTTFHEKDAKRVGLGEGKASTLDDSCQDLKRWMLGEGFALLSDSISV 180  
 Qy 124 NRSLLPDIRHPNCSKRYLETLPNTSIIIFHNEGWSSLLRTVHSVYLNRSPELVAEIVLV 183  
 Db 181 NRSVVDIHPICRKKKEYAKLPVSVIIFVNEVLSVLMRSVSHSLNRSPELMKEITLV 240  
 Qy 184 DDFSDREHLKKPLDGYMA-LFPVSAILRTKKREGILIRMLGASVATQDVTITLDSHEA 242  
 Db 241 DDHSDREYLGLELYIAEHFRKVAVVRLPRTGLIGRAAGANNATAEVLIFLDSHVEA 300  
 Qy 243 NVNMLPPLIDRIARNRKTIIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYKRIPIPE 302  
 Db 301 NVNMLPPLIEPIALNKRTAVCPFIDVIDHTNFHYAQ--DEGARQGFDMFEFFYKRLPLPE 359  
 Qy 303 LQKADSPDPSPVMAAGLFAVDRKFWELGQYDPLGELIWGEOYEYEFKVMCGGMRD 362  
 Db 360 DLK-HPAPFPKSPIMAGGLFAISREFFWELGQYDEGLIWGEOYELFKIMCGGEYD 418  
 Qy 363 IPCSRVGHIVR-----KVVPKVPAGVSLANLKVAAVMDVEYLYYQRRPE-YRLHS 416  
 Db 419 APCSRIGIYGRPNHQPSPRK--GDYLLKRYKRVAAVWMDVEYNVLYSHGDGLYESVD 475  
 Qy 417 AGDVAVQKRLSLNCKSFKFMETKIAMDLPKFYPPVPPAAAMGEIRNVT-GLCADT- 474  
 Db 476 PGDLTEQAIRTKLNCKSFKFMEEVAFDLMKTYPPVDPSPYAKALQNVGNQNLCDTL 535  
 Qy 475 ---KHGALG-----SPLRLEGCVARGEAAANNNOVFTFWREDIRPDPQHTKKF 522

Db 536 GRKGNKMGYACADNITKFOR-----TQFWELSKRDLR---LRRKKE 576  
 Qy 523 CFDA--ISHSPVLYDCHSKGNQLMKY-RKDITLYHPVG-SCMCSSESDHRIFMNTC 578  
 Db 577 CLDVQIWDANAPVWLMCHSGGQNGQYVYDRKHQKHGTGRCCLLELPPSGEVANNC 636  
 Qy 579 NPSSLTQOQWLFPHNTSSTVLEKENRN 603  
 Db 637 DTDNRFOQWNGSFNKTALDNYSD 661

## RESULT 10

Q7PZM5 PRELIMINARY; PRT; 599 AA.  
 ID Q7PZM5  
 AC Q7PZM5  
 DT 01-MAR-2004 (TREMBlrel. 26, Created)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE AGCP9340 (Fragment).  
 GN Name=agCG54449; ORFNames=ENSNANG0000012813;  
 OS Anopheles gambiae str. PEST.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
 OX NCBI\_TaxID=160454;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 CC Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AAB01008986; EAA00339.1; -.  
 DR InterPro; IPR001173; Glyco\_trans\_2.  
 DR InterPro; IPR000772; Ricin\_B\_lectin.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR Pfam; PF00652; Ricin\_B\_lectin; 3.  
 DR PROSITE; PS50231; RICIN\_B\_LECTIN; 1.  
 DR NON TER  
 SQ SEQUENCE 599 AA; 69224 MW; DA646C182B143028 CRC64;

Query Match 40.9%; Score 1341.5; DB 2; Length 599;  
 Best Local Similarity 45.8%; Pred. No. 6e-100;  
 Matches 257; Conservative 104; Mismatches 175; Indels 25; Gaps 13;  
 Qy 52 GQSHSQKQKTFIFGDOGKLDKMDKREAIIRDAQVNGEGRRPMTDAERV---DOAY 108  
 Db 46 GEGFYAMPKRVN---AGSKI-DWNNVLEIEBSKKTGEGHGRFYKLSSEODILMAKLF 100  
 Qy 109 RENGFNIYVSDKISLNRLPDIRHPNCSKRYLETLPNTSIIIFHNEGWSSLLRTVHSV 168  
 Db 101 KENYSANVSDMLNLSVDPDIRHPSCRMKEYLKLPLVSVIIFVNEHMGALLRTVYSV 160  
 Qy 169 LNRSPPELVAEIVLVDPDSREHLKKPLEDM--ALPVSAILRTKKREGILIRMLGAS 226  
 Db 161 LNRSPPELVAEIVLVNHSHTPFLMTPLREVESELAPKRVLDLPERSGILVARMGAR 220  
 Qy 227 VATGDVTITLDSHOBAAVNMPLPPLIDRIARNRKTIIVCPMIDVIDHDDFRYETQAGDAMRG 286  
 Db 221 EARGVAVLVDLSHVEVNTNMLPPLIEPIAEDYRCVCPFIDVIDHTNFQYRSQ--DEGRKG 279  
 Qy 287 AFDWMYKRIIP-IPPELQKADSPDPSPVMAAGLFAVDRKFWELGQYDPLGELIWGE 345  
 Db 280 AFDKFPYKRLPLPGDLD--DPIKFPNSPIMAGGLFAISAKFPWELGQYDEGLIWGE 337  
 Qy 346 QYEISFKVMCGGMRDIPCSRVGHIYRKYPYVPAVGS-LANNLKVAAVWMDVEYAEY 404  
 Db 338 QYEISFKIMQCGGLVDAFPCSRVGHYVRYGYPAPFGAPGVNVVNNFYKVAVWMDVEYQF 397  
 Qy 405 IYQRRPEYRHLSADVAVQKRLRSLNCKSFKFMETKIAMDLPKFYPPVPPAAAMGEIR 464  
 Db 398 LYERNPQAKTDPDLSHQRELRLRLOCKPKFWLEVAVDLIVRYPRDPOPPASGRVQ 457

```

QY 465 NVGT-G/CADTKHGALGSPLRLEGCVARGGBAANNNOVFTTWREDIRPDGPHQTKKFC 523
DB 458 SVANPRCLDLSLNQAKPEIGLYACAFNKTHP--ONNOFTLSYHRIIR---VRSNDKC 511
QY 524 FPAISHTSPVLYDCHSMKGNQMLWKY-RKDKTYLHPVS--GSCMDCESDHRIFMNTCNP 580
DB 512 LDAAKNDLDELVLFCCHESQGNQMRDYVESMTIIGHGDHRCMCEADLHTKCLFVRSCKD 571
QY 581 SSLTQOMLFHTNSTVLEKEN 601
DB 572 DKASQKWNMGVYVNFVHLQNMWD 592

RESULT 11
Q7ODRO PRELIMINARY; PRT; 1003 AA.
AC Q7ODRO;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DE ACP10665 (Fragment).
GN Name=sgCG47419; ORFNames=ENSANGC0000013497;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01008849; EAA07231.1; -.
DR InterPro: IPR001173; Glyco_trans_2.
DR InterPro: IPR000772; Ricin_B_lectin.
DR Pfam: PF00535; Glycos_transf_2; 2.
DR Pfam: PF00652; Ricin_B_lectin; 4.
DR PROSITE; PS50231; RICIN_B_LECTIN; 2.
FT NON TER
FT 1
SQ SEQUENCE 1003 AA; 115923 MW; 753EA50F567AAB13 CRC64;

Query Match 38.0%; Score 1247; DB 2; Length 1003;
Best Local Similarity 47.6%; Pred. No. 5.9e-92;
Matches 258; Conservative 77; Mismatches 179; Indels 28; Gaps 14;
QY 73 DMHDKELRRDAQVNGEGGRPYMTDARVD---QAYRANGFNIVYSKISLNSLP 128
DB 27 DYNVDEIQDNLNKGVEGQKRPYTSPEEATSELRELKLYKGFNLALSDKISINSIA 86
QY 129 DIHPNCSKRYLETTLPNTSIIIPFNHNGSSLLRTVSHVNLNRPPELVAVIVLDFSD 188
DB 87 DLNHPCKLKSYSNHLPIASVVVFYEHNSTLRTIYSLNRPPLIKETIIIVDGGST 146
QY 189 REHKKPLEDYMAL-PPSVRLRTKKEGLIRTEMLGASVATGVITFLDSHCANVNL 247
DB 147 KEFLHNKLEDYVQNLPKVKLVROPERTGLIKALAGAKIASGVLLFLDSHTAGVNL 206
QY 248 PLLLDLRARNRKTIYVCMIVDHDHDFRYEQADARGAFAWMYTKRPIPELOKAD 307
DB 207 PLLLEPRAENPKTCVCPILVIDQTDVHQ--DEGRGFLDWFHYKRVVKNK-DRIS 264
QY 308 PDPFESPVAAGLFAVDRKFMELAGYDPLIWMGEQYEISFKVMMCGRMEDICSR 367
DB 265 PTEPFPFPVWAGLFAAGADPFWELSGYDELDLWGAEOYEISFKIQCCGRMLDACS 324
QY 368 VGHYIRKTVYKVPAGVS-LARNIKRVAEVMDEYAEYIYORDE-YRHSAGDAVQKK 425
DB 325 FGHIYIRYSPFPNRSKYDFTIRNHNKRVAAIWMDEYKYIYDORDERAKTDAGMSKKT 384
QY 426 LRSLNCKSPKFMPTKIAMLPLPKYPPVPEPPAAMGRIWVG-TGLCADTKHGALGSPLR 484

```

```

DB 385 IRETLMKPFRKMFLOEFAPEILIEYFVEPEPVASSGSLQSVADSLCIDTMQRGRGEIG 444
QY 485 LEGCVARGGBAANNNOVFTTWREDIRPDGPHQTKKFC--CPDAISHT--SPVLYDCH 539
DB 445 LYPSCNSLIEPT-NHNGYFVHSHWRI-----QH-KYGECCFDPVQSPKSPVTLIFTC 496
QY 540 SMKGNQMLKTKKTYLHPVSGCMD-----CSESDHRTFMNTCNPSSLTQOMLFHTNST 595
DB 497 MHQGNQFPQY-DHKTQQIKNGVCIDSDPAKVECHRINYSKULPQTSIIVFPDEHWST 555
QY 596 VL 597
DB 556 LL 557

RESULT 12
GLT4 DROME STANDARD; PRT; 659 AA.
AC Q81A42; Q81Q11;
DT 01-OCT-2004 (Rel. 45, Created)
DT 01-OCT-2004 (Rel. 45, Last sequence update)
DE 01-OCT-2004 (Rel. 45, Last annotation update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE N-acetylglucosaminyltransferase 4 (EC 2.4.1.-) (Protein-UDP
DE acetylglucosaminyltransferase 4) (UDP-GalNAc:polypeptide N-
DE acetylglucosaminyltransferase 4) (pp-GalNAse 4).
GN Name=pgant4; ORFNames=CG31955;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., ENZYME ACTIVITY, TISSUE SPECIFICITY, AND
RP DEVELOPMENTAL STAGE.
RC STRAIN=Canton-S; TISSUE=Embryo;
RX MEDLINE=22841110; PubMed=12839714; DOI=10.1074/jbc.M3036200;
RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.,
RT "Functional characterization and expression analysis of members of the
RT UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase family from
RT Drosophila melanogaster."
RL J. Biol. Chem. 278:35039-35048 (2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballow R.M., Baer A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brocklehan P., Brothier P.,
RA Butts K.C., Busam D.A., Butler H., Cadietu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Maye A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Feiler K.C., Gabriellian A.B., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodde A.L., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nushekn D.R., Pacלב J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```





```

QY 506 TWREDIRPGDPOHTKKE---CFDAISHTSP-----VTLYDCHSNKGNQMLMKY-RKDKTLVH 557
D 557 S1RDLR-----MKGFDVCLDV--HEGPNATVMMWSCHSGGNGFWYTDROTQLVH 608
QY 558 PVGS-CMD--CSESDHRIFMNTCNPSLSLTOQWLFHTNSTVLEKEF 600
D 609 GENNRCLCEGVENGIAKVAVANSCEGNDQRQWRFEGVNTHTMLDTF 654

RESULT 13
AA056701 PRELIMINARY; PRT; 659 AA.
ID AA056701
AC AA056701;
DT 02-MAR-2004 (TREMUREL. 27, Created)
DT 02-MAR-2004 (TREMUREL. 27, Last sequence update)
DT 02-MAR-2004 (TREMUREL. 27, Last annotation update)
DE UDP-GalNac:polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41).
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=1829714;
RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.;
RT "Functional Characterization and Expression Analysis of Members of the
RT UDP-GalNac:polypeptide N-acetylglactosaminyltransferase Family from
RT Drosophila melanogaster."
RL J. Biol. Chem. 278:35039-35048(2003).
DR EMBL; AY268065; AA056701.1; -
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 659 AA; 75805 MW; 0182D42ABAA1DD2 CRC64;

Query Match 36.2%; Score 1186; DB 2; Length 659;
Best Local Similarity 41.2%; Pred. No. 87;
Matches 266; Conservative 106; Mismatches 216; Indels 58; Gaps 21;

QY 2 RRKEKRLQAVLALVLLPNVGLMALYR-----EROPDG---TPGSGAAYA 48
D 20 KRYVKKLRKRVLLVIVVSVLTIVVERMKMAALETRQDLPNDPITVPRANIH 79
QY 49 PA--ACGQSHSRQKTFILDDGQTL-----KDWKKAIRRDQAVNGEQR 94
D 80 PTRKAPRPFPQDRSVVDIPRSDLGQFRLPEPKERKMDWDYAAWADRKRSGFGEHV 139
QY 95 PYRM--TDARVQ--QAVRENGFNIVYDSTLSRSLDIPRNCNSKRYLETLPNTSITI 151
D 140 AVKIEPDERQOLEKEHEENMGPNLISDRISVNSVBDLRLEACKTKRYLAKLPNTSVIF 199
QY 152 PFNHEGSSILRTVSVLNSRSPPELVAVIYLVDDPSREHLKKPLBYMAL-FPS-VRIL 209
D 200 IFNNEHNTILKRSYVYNTTPPELLAQIYLVDDGSEMDVLKQGLDYVQOHPHLVTV 259
QY 210 RTKREGLITRMLGASVATGDTITFLDSHCANVMNLPILLRIARNRRTIVCPMDIVI 269
D 260 RNDERQGLIGARLAGAVAGVAVGVWVFPDISHIEVYVNLPLILBIAINPKISTCPMDVTI 319
QY 270 DHDDFERETQAGAMRAFPWEMTYKRIPIPELQKADPSDPSPMAGGLFVAVDKWF 329
D 320 SHEDFSYFSGNKGARGGPFMKMLYKQLPYLPB-DALDKMPYSPVWVGGLFINTDF 378
QY 330 WEIGGYDPLGIEINGEYOEISFPKVMCGRMEDIPSGVGIYR-KVVPKVPAGVS-LA 387
D 379 WDLGSDIDDDIDINGEYOELSFKIMCGMLLDVPCRVANITFGPKPKPKNPNFHNFA 438
QY 388 RNLKRVAVWVWDEAYEYIYORRPE-YRHLGADVAVOKLRSLSLNCKSFKMFTKIAML 446
D 439 KNRKRVAVWVWDEAYEYIYORRPE-YRHLGADVAVOKLRSLSLNCKSFKMFTKIAML 498

```

```

QY 447 KRYPPVPPAAAMGEIRNVGTGL-CADTKGALGSLRLEGYGRGGAAMNMOYFTF 505
D 499 LVKFPVPPSYAAGIIQVANVYVCLDNKNGKSTEEAVGCFSCADNRTHQPN--QWEL 556
QY 506 TWREDIRPGDPOHTKKE---CFDAISHTSP-----VTLYDCHSNKGNQMLMKY-RKDKTLVH 557
D 557 S1RDLR-----MKGFDVCLDV--HEGPNATVMMWSCHSGGNGFWYTDROTQLVH 608
QY 558 PVGS-CMD--CSESDHRIFMNTCNPSLSLTOQWLFHTNSTVLEKEF 600
D 609 GENNRCLCEGVENGIAKVAVANSCEGNDQRQWRFEGVNTHTMLDTF 654

RESULT 14
ID GLT9_DROME STANDARD; PRT; 650 AA.
ID GLT9_DROME
AC Q8MRG9; Q9V770;
DT 01-OCT-2004 (Rel. 45, Created)
DT 01-OCT-2004 (Rel. 45, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Putative polypeptide N-acetylglactosaminyltransferase 9 (EC 2.4.1.41)
DE (Protein-UDP acetylglactosaminyltransferase 9) (UDP-
DE GalNac:polypeptide N-acetylglactosaminyltransferase 9) (p-GANTase
DE 9).
GN Name=pGant9; ORFNames=CG30463;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Ahril J.F., Agbayani A., An H.-D., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson B., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hseltin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasok P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris D.M., Nelson D.L.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skyski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Syntreks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasmann D.A., Weinstock G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zavari J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]

```



RP		REVIEWSIONS.
RX	MEDLINE=22426069; PubMed=12537572;	
RA	Mitra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,	
RA	Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochank S.E.,	
RA	Smith C.D., Tapp J.L., Whitfield E.J., Bayraktaroglu I., Bernan B.P.,	
RA	Bettencourt B.R., Celtniker S.E., de Grey A.D.N.J., Drysdale R.A.,	
RA	Harris N.L., Richter C., Russo S., Schroeder A.J., Shu S.Q.,	
RA	Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,	
RA	Lewis S.E.;	
RT	"Annotation of the Drosophila melanogaster euchromatic genome: a	
RT	systematic review."	
RL	Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Berkely; TISSUE=Embryo;	
RX	MEDLINE=22426066; PubMed=12537569;	
RA	Stapleton M., Carlson J.W., Brokerstein P., Yu C., Champagne M.,	
RA	George R.A., Guanin H., Kronmiller B., Pacleby J.M., Park S., Wan K.H.,	
RA	Rubin G.M., Celtniker S.E.;	
RT	"A Drosophila full-length cDNA resource."	
RL	Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).	
CC	-I- FUNCTION: May catalyze the initial reaction in O-linked	
CC	oligosaccharide biosynthesis, the transfer of an N-acetyl-D-	
CC	galactosamine residue to a serine or threonine residue on the	
CC	protein receptor (By similarity).	
CC	-I- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =	
CC	UDP + N-acetyl-D-galactosaminyl-polypeptide.	
CC	-I- COFACTOR: Manganese and calcium (By similarity).	
CC	-I- PATHWAY: Glycosylation.	
CC	-I- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By	
CC	similarity).	
CC	-I- DOMAIN: There are two conserved domains in the glycosyltransferase	
CC	region: the N-terminal domain (domain A, also called GT1 motif),	
CC	which is probably involved in manganese coordination and substrate	
CC	binding and the C-terminal domain (domain B, also called	
CC	Gal/Galnac-T motif), which is probably involved in catalytic	
CC	reaction and UDP-Gal binding (By similarity).	
CC	-I- DOMAIN: The ricin B-type lectin domain binds to Galnac and	
CC	contributes to the glycopeptide specificity (By similarity).	
CC	-I- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNac-T	
CC	subfamily.	
CC	-I- SIMILARITY: Contains 1 ricin B-type lectin domain.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AE003806; AAF57964.2; --	
DR	EMBL; AY121661; AAAS1988.1; --	
DR	HSSP; P26514; IKM.	
DR	FlyBase; FBgn0050463; CG30463.	
DR	InterPro; IPRO01173; Glyco_trans_2.	
DR	InterPro; IPRO09061; Putative DNA_bind.	
DR	InterPro; IPRO08997; RicinB_Like	
DR	InterPro; IPRO00772; Ricin_B_lectin.	
DR	Pfam; PF00535; GlycoS_transf_2.1.	
DR	Pfam; PF00652; Ricin_B_lectin; 3.	
DR	SMART; SMO0458; RICIN; 1.	
DR	PROSITE; PS50231; RICIN_B_LECTIN; 1.	
KW	Calcium; Glycosyltransferase; Golgi stack; Lectin; Manganese;	
KM	Signal-anchor; Transferase; Transmembrane.	
FT	DOMAIN 1 11	Cytoplasmic (Potential).
FT	TRANSMEM 12 31	Signal-anchor for type II membrane
FT		proteins (Potential).
FT	DOMAIN 32 50	luminal (Potential).
FT	DOMAIN 308 317	Catalytic subdomain A.
FT	DOMAIN 378 440	Catalytic subdomain B.
FT	DOMAIN 521 643	Ricin B-type lectin.
FT	DOMAIN 43 90	Gly-rich.

Query Match	Beet Local Match	Similarity	35.8% 40.2% 272; Conservative	Score Pred. No. 3.6e-86; 91; Mismatches 201; Indels 113; Gaps	DB 1; Length 650; Indels 113; Gaps
FT	DISTLFLD	535	554	By similarity.	
FT	DISTLFLD	577	590	By similarity.	
FT	DISTLFLD	616	631	By similarity.	
FT	CAROHND	321	323	N-linked (GlcNAc. . .) (Potential).	
FT	CAROHND	373	373	N-linked (GlcNAc. . .) (Potential).	
FT	CONFLICT	145	145	N -> D (in Ref. 3).	
FT	CONFLICT	454	454	L -> P (in Ref. 3).	
FT	CONFLICT	472	472	Y -> C (in Ref. 3).	
SEQ	SEQUENCE	650 AA;	73192 MW;	AC847736AD1C07CA CRC64;	
Query Match	Beet Local Match	Similarity	35.8% 40.2% 272; Conservative	Score Pred. No. 3.6e-86; 91; Mismatches 201; Indels 113; Gaps	DB 1; Length 650; Indels 113; Gaps
QY	2 RREKRELLQAVALVLAALVLLPNNGLMALYERPDGTGGSGAVALPAAGQS-----	55			
DB	6 RRRSTTYKLVAFALAIWFCIAFLVYTDTRRRAAQEGASGASGASAGVGAGGLDP	65			
QY	56 -----HSRÖKKTFPL-----GDSÖKLKMDHD-----	76			
DB	66 IALALRNPAEEDFGINGNVIGGGQ--KQAHDEADIPPTVGKHKADLQASRMKKAQ	123			
QY	77 -KEAIRPAAGV-----GNGEQGRPY-----PMTD--AEYDQAVREGNFYVSDK	120			
DB	124 PKGRQESQSKYIDPDPANFEENPGELGVRVLPKEMSEMKKAVDGGTKAFLNQVSDL	183			
QY	121 ISLNRSLDPIRHPNCSNK-RYLETLTPNLSIIIPFNBSGMSILRTVSHVLSRSPETVAE	179			
DB	184 ISVHTLPDPDAMCKDEARVLTNLPKTDVLIICFNEMAYTLRTVSHVLSRSPHLLGK	243			
QY	180 IIVVDSEDRHLKKPLEDYVALPFSVRLTLTKRREGILIRPMIGASVATGDTVTFLDH	239			
DB	244 IIVVDYSDMHLTKQLEDYPAAYPKVQIIIRGQKEGILIRARIIIGANNHAKSPVLYTLDH	303			
QY	240 CEANVMPLPDLDRARARAKTIYCPMDIVIDHDPRYE--TQAGDAMRGAPFMEWYKXIP	298			
DB	304 CECTEGMLPELLDRARAKSTTVCPVIDVDSDELTLEYHRDSSGGVNVGCFPMNLQFSWH	363			
QY	299 IPEPLQKADPS--DPFESPVAAGGLFAVDRKMFELGSDYDQGLTIWGEQYEISPKVMC	356			
DB	364 VPERRKGNHNTAEAEVYSPMAAGGFSIDREFPDLGLTYDGGFDIWSGENLELSKTMWC	423			
QY	357 GGRMEDIPCSRAGHLYRKTYVPYKQPAVS-LARNLKRAEYVMDYAEIYIQR---RPE	411			
DB	424 GGTLEIYVCSHVGHIFRKSPYKMSGVNVLKJNSVRLAEVWMDYSEGYHYHRIIGNDKD	483			
QY	412 YRLHSLAGVAOYKTLRSLNCSKSPKFWTKLAMDLPKRYPEVERP--AAAGCEIRNVGTG	469			
DB	484 W-----GVSSRRKRLRNDLKCSKPFMYLDNI-----YPFLFIIPGSDVAAGEIRNLIDY	531			
QY	470 --LCADTGHG-----ALGSPRLBEGCVRGGEAAMNNMQVFTFWREDIRPGDPHTKC	521			
DB	532 GRTCLDAPAGKKHQAQVGT-----YPCIRGGNGYV-----MLSKAGEIRRDS-----	576			
QY	522 FCFDAISHTSPVLYDCHSMKGNOLMKYRKD-KTLVHPVSGSCMDSSDHRIFMNTCNP	580			
DB	577 -CLDYAG--KDYTLFGCHGKGNGWTYRENTKQLHGTSGKCLAISSSKDKLMEBCSA	633			
QY	581 SSLVYQMLPEHTNSTVL	597			
DB	634 SLRQOWTLENYDSKL	650			
RESULT 15					
ID	Q70815	PRELIMINARY	PRT;	645 AA.	
AC	Q70815;				
DT	01-MAR-2004 (Tremblrel. 26, Created)				
DT	01-MAR-2004 (Tremblrel. 26, Last sequence update)				
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)				
DE	Eblp5044 (fragment).				
OS	Name=eblG5044; ORFNames=ENGANG0000003900;				
OS	Anopheles gambiæ str. PEST.				

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anophel.  
 OC NCBI\_TaxID=180454;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 DR EMBL; AA001008944; EMBL0180.1; -;  
 DR InterPro: IPR001173; Glyco\_trans\_2;  
 DR InterPro: IPR000772; Ricin\_B\_lectin.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR Pfam; PF00652; Ricin\_B\_lectin; 3.  
 DR PROSITE; PSS0231; RICIN\_B\_LECTIN; 1.  
 FT NON TER 645  
 FT SEQUENCE 645 AA; 73459 MW; AADPDB6CE16B51C0 CRC64;

Query Match 35.4%; Score 1159; DB 2; Length 645;  
 Best Local Similarity 44.3%; Pred. No. 4.4e-85;  
 Matches 251; Conservative 79; Mismatches 193; Indels 44; Gaps 19;  
 QY 38 GTGGGGAATAAPAGGSGSHRQKTFPLGSGQKXMDHDEKATIRDAQVGN-GEGRP- 95  
 DB 101 GDEPGYGG--GGGRADSSMPRTY--RPELKKRRQAPTV--KENYGRPEWGPV 151  
 QY 96 -YPMTDARYDQAVRENGFNIVYSDKISLNRSLPDIRHPNCNSKRYLETPTNTSIIIPFH 154  
 DB 152 KIRANQGEMLKEKEKQENQFPLASDMWLRSLTDVHHQCKKHAYAKLPTTSIVVPH 211  
 QY 155 NEGWSSILRTVHSVYLNRPPELVAEIVLVDPSDRHKKPLEDYMALFP-SVRIITPKK 213  
 DB 212 NEAMSTLRTIWSYINRSPRLKEIILVDASERHGGQLEBYVTLTPVPTVLTGK 271  
 QY 214 REGIIRRMIGASATGDTVTFLDSHCEAVNMLPRLDRIARRKTIYCPMTDIVDHD 273  
 DB 272 RSGIIRRLGAKKVKQGVITFLDACECTEGMELPLARIVLDRKTVCPITIDISDET 331  
 QY 274 FRYETQAGDAMRGAFDWMETTKRIPPE-ELQKA--DPSDPFESPVNAGGLFAVDRKMF 330  
 DB 332 FEVYAT-ASDQTWGFNMKLMFRWRYVARERQNRNHTAPLRPTMAAGLFSIDRYFY 390  
 QY 331 ELGQYDPLGIEWGEOYEISFKVMCGRMEDIPCSFVGHITRYKYVPYKVPAGVS--LAR 388  
 DB 391 EISYDSDGMDIMGGENLEMSFRVVMCGTLEIAPCSRVGHVFRKSTYSPFGGTSIVNK 450  
 QY 389 NLRVAEVMDEVAEYIYQRRPEFRHSAGDVAVQKLRSLNCKSKFRMPTKIAMLPK 448  
 DB 451 NNARLAEVMDGWSSEFYTNINPGARKASADVSERRALRELRCKSKFRMYLENI----- 504  
 QY 449 FYFPEVPEPALAM--GEIRNVGTGCAPTDKGALGSPRLBEGCVAGREBAAMNNQVTF 506  
 DB 505 -YESQMPLYIYFGEIRNVKTNCLDITGRKSKNEKIGSSYC-HGLG-----GNQVAYT 557  
 QY 507 WRDIRGDPQHTKKCFDAISHTSPVTLTYDCSHMKGNOLMKY-RKDKTLVHPVSGSCMD 565  
 DB 558 KRHQIMSDDN-----CLDASNALGPVNLVRCHMGNGNOEWIYDDEKTIKRVNSGKCLT 611  
 QY 566 -CESDHRIT-FMNTCNSSSLTQWTFE 590  
 DB 612 RASEDSDPTPLRCPNYS-QQQWLMQ 637

## RESULT 16

ID Q7Q048 PRELIMINARY; PRT; 518 AA.  
 AC Q7Q048;  
 DT 01-MAR-2004 (Tremblrel. 26, Created)  
 DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)  
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
 DE AgCP9928 (Fragment).

GN Name=agCG55772; ORFNames=ENSNAGG00000013981;  
 OS Anopheles gambiae str. PEST.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anophel.  
 OC NCBI\_TaxID=180454;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 DR EMBL; AA001008986; EMBL0233.1; -;  
 DR InterPro: IPR001173; Glyco\_trans\_2;  
 DR InterPro: IPR000772; Ricin\_B\_lectin.  
 DR Pfam; PF00535; Glycos\_transf\_2; 1.  
 DR Pfam; PF00652; Ricin\_B\_lectin; 3.  
 DR PROSITE; PSS0231; RICIN\_B\_LECTIN; 1.  
 FT NON TER 1  
 FT SEQUENCE 518 AA; 58964 MW; 00F8821266B395FE CRC64;

Query Match 34.3%; Score 1144; DB 2; Length 518;  
 Best Local Similarity 45.7%; Pred. No. 5.4e-84;  
 Matches 243; Conservative 71; Mismatches 170; Indels 48; Gaps 15;  
 QY 88 GNGEGRPEYMTD-----AEVDAQRENGFNIVYSDKISLNRSLPDIRHPNCNS-KRY 140  
 DB 13 GPELGRFPVLPPELPEYVKLVDDGAKNAFQVYADMISIRRTLPDPDAWCKEPPRY 72  
 QY 141 LETLPTNTSIIIPPHNEGSSSLRTVHSVYLNRPPELVAEIVLVDPSDRHKKPLEDY 200  
 DB 73 MDLPPTSVLIICFHNEMASVLTIRVHSVLDSPQVLEKIVLVDPSDMPTQKEDYF 132  
 QY 201 ALPESVYILRTKKREGIIRMTGASVATGDTVTFLDSHCEAVNMLPRLDRIARRK 260  
 DB 133 LAIPKIKIYAAARBEIIRARLLGARHATAPVLTIDSHCECYGMELPLDRIARSTT 192  
 QY 261 IYCPMTDIVDHDFFRYE-TQAGDAMRGAFDWMETTKRIPPE-ELQKA-DPSDPFES 317  
 DB 193 VCPVVIDVIDDNTMEYHYRSDGCVNNGVGFDMNLQFMHNAVBEKRGKSPAEVWSP 252  
 QY 318 AGGLFANDRKMFHELGYPDGLIEWGEOYEISFKVMCGRMEDIPCSFVGHITRYKY 377  
 DB 253 AGGIFALDRVFFRLGTGDFIDGGENLEISFKTMCGSLEIYPCSHVGHIFRRSP 312  
 QY 378 YKVPAGVS-LARLKRVAEVMDEVAEYIYQRRPEFRHSAGDVAVQKLRSLNCKSK 432  
 DB 313 YKRTGVNVTKRNSVRLAEVMDGWSSEFYTNINPGARKASADVSERRALRELRCK 367  
 QY 433 KSFKEWPTKIAMDLPKFYFPEVPEPP--AAWGEIRNVGTG--LCADTKGA--LGSPLRL 486  
 DB 368 KSFRTYLDNI-----YPLFPVPGDAVASGEVRNNGYGRITCLDAPGKRLRKPVGLY 420  
 QY 487 GCYRGRGEAAMNNQVTFTFWRDIRGDPQHTKKCFDAISHTSPVTLTYDCSHMKGN 546  
 DB 421 PCNRQGNQYM--MLSKTGEIRD-----EACLDVAG--DDVVLTPCHGSRGNQY 466  
 QY 547 WKTR-KDKTLVHPVSGSCMDSESDHRIFMNTCNPSLSLTQWTFEBHTNSTVL 597  
 DB 467 WNTSDGSHLRHSSDRCLAINAKNKLIMQDCAVAEQRWSPQNTDAKL 518

## RESULT 17

ID GLTI\_HUMAN STANDARD; PRT; 559 AA.  
 AC Q10472; Q9D086;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Polypeptide N-acetylglucosaminyltransferase 1 (EC 2.4.1.41)  
 DE (protein-UDP acetylglucosaminyltransferase 1) (UDP-  
 DE GalNAc:polypeptide N-acetylglucosaminyltransferase 1) (Polypeptide

DE GalNAc transferase 1) (GalNAc-T1) (pp-Gantase 1) [Contains:  
 DE Polypeptide N-acetylgalactosaminyltransferase 1, soluble form].  
 GN Name=GalNT1;  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Salivary gland;  
 RX MEDLINE=96115928; PubMed=6690719;  
 RA Meurer J.A., Naylor J.M., Baker C.A., Thomsen D.R., Homa F.L.,  
 RA Elhammer A.P.;  
 RT "CDNA cloning, expression, and chromosomal localization of a human  
 RT UDP-GalNAc:polypeptide, N-acetylgalactosaminyltransferase.";  
 RL J. Biochem. 118:568-574 (1995).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.  
 RX MEDLINE=96025800; PubMed=7592619;  
 RA White T., Bennett E.P., Takio K., Soerensen T., Bonding N.,  
 RA Clausen H.;  
 RT "Purification and cDNA cloning of a human UDP-N-acetyl-alpha-D-  
 RT galactosamine:polypeptide N-acetylgalactosaminyltransferase.";  
 RL J. Biol. Chem. 270:24156-24165 (1995).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=PNS;  
 RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins P.S., Wagner L., Shemmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA DiCicco L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uebin T.B., Toshitsuki S., Carrinci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton B., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez Y.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska J., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [4]  
 RP SEQUENCE OF 512-559 FROM N.A. (ISOFORM 1).  
 RX MEDLINE=96285863; PubMed=8727794;  
 RA Meurer J.A., Drong R.F., Homa F.L., Slightom J.L., Elhammer A.P.;  
 RT "O-galactosylation of a human UDP-GalNAc:polypeptide, N-  
 RT acetylgalactosaminyltransferase gene and a related processed  
 RT pseudogene.";  
 RL Glycobiology 6:231-241 (1996).  
 RN [5]  
 RP FUNCTION.  
 RX PubMed=9295285;  
 RA Wandall H.H., Haasan H., Mirgorodskaya E., Kristensen A.K.,  
 RA Roepstorff P., Bennett E.P., Nielsen P.A., Hollingsworth M.A.,  
 RA Burchell J., Taylor-Papadimitriou J., Clausen H.;  
 RT "Substrate specificities of three members of the human UDP-N-acetyl-  
 RT alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase  
 RT family, GalNAc-T1, -T2 and -T3.";  
 RL J. Biol. Chem. 272:23503-23514 (1997).  
 RN [6]  
 RP SUBCELLULAR LOCATION.  
 RX PubMed=9394011;  
 RA Roepstorff S., White J., Wandall H.H., Olivo J.-C., Stark A.,  
 RA Bennett E.P., Whitehouse C., Berger E.G., Clausen H., Nilsson T.;  
 RT "Localization of three human polypeptide GalNAc-transferases in HeLa  
 RT cells suggests initiation of O-linked glycosylation throughout the

RT Golgi apparatus.";  
 RL J. Cell Sci. 111:45-60 (1998).  
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has a broad spectrum of substrates for peptides  
 CC such as EA2, Muc5AC, Muc6, Muc7, Muc8, Muc9, Muc10, Muc11,  
 CC CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium.  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi; resides  
 CC evenly across the Golgi stack. A secreted form also exists.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named Isoforms=2;  
 CC Name=1;  
 CC IsoId=Q10472-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q10472-2; Sequence=VSP\_011200;  
 CC Note=No experimental confirmation available;  
 CC -1- TISSUE SPECIFICITY: Widely expressed. Expressed in all tissues  
 CC tested.  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called GT1 motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC Gal/GalNAc-T motif), which is probably involved in catalytic  
 CC reaction and UDP-gal binding (By similarity).  
 CC -1- DOMAIN: The ricin B-type lectin directs the glycopeptide  
 CC specificity. It is required in the glycopeptide specificity of  
 CC enzyme activity but not for activity with naked peptide  
 CC substrates, suggesting that it triggers the catalytic domain to  
 CC act on GalNAc-glycopeptide substrates (By similarity).  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
 CC subfamily.  
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U4514; AAC50327.1; -;  
 CC EMBL: X85019; CA59380.1; -;  
 CC EMBL: BC047746; AA47746.1; -;  
 CC EMBL: S82597; AAD14406.1; -;  
 CC PIR: JC4223; JC4223.  
 CC GeneW: HGNC:4123; GALNT1.  
 CC MW: 602273; -;  
 CC GO: GO:0004653; F:polypeptide N-acetylgalactosaminyltransferase. . .; TAS.  
 CC GO: GO:0006493; P:O-linked glycosylation; TAS.  
 CC InterPro: IPR001173; Glyco\_trans\_2.  
 CC InterPro: IPR008997; Glyco\_1like.  
 CC InterPro: IPR000772; Ricin\_B\_lectin.  
 CC Pfam: PF00535; Glycos\_transf\_2; 1.  
 CC Pfam: PF00652; Ricin\_B\_lectin; 3.  
 CC SMART: SM00458; RICIN\_1.  
 CC PROSITE: PS50231; RICIN\_B\_LECTIN; 1.  
 CC Alternative splicing; Calcium; Glycoprotein; Glycosyltransferase;  
 CC Golgi stack; Lectin; Manganese; Signal-anchor; Transferase;  
 CC Transmembrane.  
 CC PROPEP 1 40  
 CC Removed in soluble polypeptide N-  
 CC acetylgalactosaminyltransferase (By  
 CC similarity).  
 CC CHAIN 41 559  
 CC Polypeptide N-  
 CC acetylgalactosaminyltransferase 1,  
 CC soluble form.  
 CC DOMAIN 1 8  
 CC Cytoplasmic (potential).  
 CC Signal-anchor for type II membrane  
 CC protein (potential).  
 CC TRANSMEM 9 28

FT	DOMAIN	29	559		Luminal, catalytic (potential).
FT	DOMAIN	115	225		Catalytic subdomain A.
FT	DOMAIN	285	347		Catalytic subdomain B.
FT	DOMAIN	429	551		Rich B-type lectin.
FT	SITE	141	141		Not glycosylated (By similarity).
FT	DISULFID	442	459		By similarity.
FT	DISULFID	482	497		By similarity.
FT	DISULFID	523	540		By similarity.
FT	CARBOHYD	95	95		N-linked (GlcNAc. . .) (By similarity).
FT	CARBOHYD	117	117		N-linked (potential).
FT	CARBOHYD	118	118		O-linked (potential).
FT	CARBOHYD	119	119		O-linked (potential).
FT	CARBOHYD	288	288		O-linked (potential).
FT	CARBOHYD	541	541		N-linked (GlcNAc. . .) (potential).
FT	CARBOHYD	552	552		N-linked (GlcNAc. . .) (By similarity).
FT	VARSPLIC	106	559		Missing (in isoform 2).
FT					/FTID=VSP 011200.
SEQ	SEQUENCE	559 AA;	64219 MM;	CD68118CB021BESB CXC64;	
	Query Match	34.3%;	Score 1125; DB 1;	Length 559;	
	Best Local Similarity	44.0%;	Pred. No. 2,1e-82;		
	Matches 232; Conservative	87;	Mismatches 170;	Indels 38; Gaps 16	
Qy	GNGEQGRP--YPMTDAERVDAYRNGNGNIVYSDKISLNRSPLDRIHNGNSKRYLELP	145			
Db	57 GPGEGKRVVYIPKEDQEKKEKFKINQFNILASENIALNRSIPVRLSGCKTKVPPDMLP	116			
Qy	146 NTSIIIPFNHSGWSSLRTVHSLVLSRSPPELVAELIVLDVDFSDRHLKKPLEDYV-ALFP	204			
Db	117 TTSVIVVHNEMASLTLRTVHSLVINSRSHRIIEELVLDVDDASERFLKRPLESYVKLV	176			
Qy	205 SVRLIRTKKRGELIRTRMLGASVATGDIYTFILDSHCANVMVLPPLDLRIANRRTIYCP	264			
Db	177 PVHIVRMEQRSGLIFARLKGAASVSKQYITFLDAHCECTVGLBELLARIKHDRRTVVC	236			
Qy	265 MIDVIYDHDPEFYEQAG-DARFGADMEYTKRIIP---ELQKADSDSPESVVMAG	320			
Db	237 IIVDISDITFEY--MAGSDMTYGGNMKLNPFMYVVPREMDRRRGDRTLVPRTVMAGG	294			
Qy	321 LFAVDRKFMFELGCDYDPLGLIETWGEQYEISFVNMVCGGREDIPCSRGHIVRYKVPYKV	380			
Db	295 LFSIDRDVFOEIGTIDYDAGMDINGENLEISPIWCGGTLELVYTSYHGHVRYKATPYTP	354			
Qy	381 PAGVS--LARNLKRVAYVMDEYAEIYIYORREYRHLASAGDVAVQKLRSSINCKSPKMF	438			
Db	439 MTKIAMD--LKFYIPPVPEPPAAWGEIRNNGTGLCADTRHGALSGPLRLBGCVRGRGEAA	496			
Qy	415 LENTYPPDQIIPRHY-----FSLDEIRNVEVNOQLDMMARKENKEVGI FNC -HGMG---	463			
Qy	497 WNNQVPEFTFRREDIRPGDPQHTKFCFCDPAISHTSPVTLIDCHSKKGNQWLYKRDCK--TL	555			
Db	464 --GNQVFSYTKANKERLRTD-----LCIDVSKLNGPVTMLKCHHLKNGQWMEYDEVKLT	515			
Qy	556 YHPVSGSCMD-CSESDHRI-FMNTGNPESLITQOMLFETHNSTVLEKEF	600			
Db	516 QHVNNSNOCLDKATEEDSQVPSIRDCN-GSRSGQWLRL--NVTLPPIIF	559			
RESULT 18					
GLTI_PIG	ID	GLTI_PIG	STANDARD;	PRT;	559 AA.
AC	Q29121;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	01-OCT-2004 (Rel. 45, Last annotation update)				
DB	Polypeptide N-acetylglactosaminyltransferase 1 (EC 2.4.1.41)				
DB	(Protein-UDP acetylglactosaminyltransferase 1) (UDP-				
DB	GalNAc:polypeptide N-acetylglactosaminyltransferase 1) (Polypeptide				
DB	GalNAc transferase 1) (GalNAc-T1) (pp-GalTnase 1) [contains:				
DB	Polypeptide N-acetylglactosaminyltransferase 1, soluble form].				
DB	Name=GALNT1;				

CC	Sus scrofa [Pig].
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX	NCSI_taxonomy9823;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=lung;
EX	MEDLINE=66318021; PubMed=8748160;
RA	Yoshida A., Hara T., Ikenaga H., Takeuchi M.,
RT	"Cloning and expression of a porcine UDP-galNAc: polypeptide N-
RL	acetylglucosaminyl transferase.";
RU	Glycoconj. J. 12:824-828(1995).
CC	-I- FUNCTION: Catalyzes the initial reaction in O-linked
CC	oligosaccharide biosynthesis, the transfer of an N-acetyl-D-
CC	galactosamine residue to a serine or threonine residue on the
CC	protein receptor. Has a broad spectrum of substrates for peptides
CC	such as EA2, Muc5AC, Muc4, Muc6 and Muc7 (By similarity).
CC	-I- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
CC	UDP + N-acetyl-D-galactosaminyl-polypetide.
CC	-I- COFACTOR: Manganese and calcium (By similarity).
CC	-I- PATHWAY: Glycosylation.
CC	-I- SUBCELLULAR LOCATION: Type II membrane protein. Golgi; resides
CC	evenly across the Golgi stack. A secreted form also exists (By
CC	similarity).
CC	-I- DOMAIN: There are two conserved domains in the glycosyltransferase
CC	region: the N-terminal domain (domain A, also called GRI motif),
CC	which is probably involved in manganese coordination and substrate
CC	binding and the C-terminal domain (domain B, also called
CC	Gal/galNAc-T motif), which is probably involved in catalytic
CC	reaction and UDP-Gal binding (By similarity).
CC	-I- DOMAIN: The ricin B-type lectin domain directs the glycopeptide
CC	specificity. It is required in the glycopeptide specificity of
CC	enzyme activity but not for activity with naked peptide
CC	substrates, suggesting that it triggers the catalytic domain to
CC	act on galNAc-glycopeptide substrates (By similarity).
CC	-I- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
CC	subfamily.
CC	-I- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/isb- ent send an email to license@isb-sib.ch). -----
DR	EMBL; D85389; BAA12800.1; -
DR	InterPro; IPRO01173; Glyco_trans_2.
DR	InterPro; IPRO08997; RicinB_1like.
DR	InterPro; IPRO00772; Ricin_B_lectin.
DR	Pfam; PF00535; Glycos_transf_2; 1.
DR	Pfam; PF00662; Ricin_B_lectin; 3.
DR	SMART; SM00458; RICIN; 1.
DR	PROSITE; PS00231; RICIN_B_LECTIN; 1.
KW	Calcium; P550231; Glycosyltransferase; Golgi stack; Lectin;
KW	Manganese; Signal-anchor; Transmembrane.
FT	PROPEP 1 40
FT	Removed in soluble polypeptide N-
FT	acetylglucosaminyltransferase (By
FT	similarity).
FT	CHAIN 41 559
FT	Polypeptide N-
FT	acetylglucosaminyltransferase 1,
FT	soluble form (By similarity).
FT	Cytoplasmic (Potential).
FT	TRANSHEM 9 28
FT	Signal-anchor for type II membrane
FT	protein (Potential).
FT	Lumenal, catalytic (Potential).
FT	Catalytic subdomain A.
FT	Catalytic subdomain B.
FT	Ricin B-type lectin.
FT	SITE 141 141
FT	Not glycosylated (Probable).
FT	DISULFID 442 459
FT	By similarity.
FT	By similarity.

FT DISULFID 523 540 By similarity.  
 FT CARBOHYD 95 95 N-linked (GlcNAc...) (By similarity).  
 FT CARBOHYD 117 117 O-linked (potential).  
 FT CARBOHYD 118 118 O-linked (potential).  
 FT CARBOHYD 119 119 O-linked (potential).  
 FT CARBOHYD 288 288 O-linked (potential).  
 FT CARBOHYD 552 552 N-linked (GlcNAc...) (By similarity).  
 SQ SEQUENCE 559 AA; 64118 MW; 6802261A516846C CRC64;  
 Query Match 34.1%; Score 1117; DB 1; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 9.4e-82;  
 Matches 230; Conservative 90; Mismatches 169; Indels 38; Gaps 16;  
 88 GNGEGRRP--YPMIDAEKVDQAYRNGENIYVSDKISLNRSLPDRHPNCSKRYLETLP 145  
 57 GPGEMKGVVPEKEDQCKMKFNQFNMASEMIALNRSLPDRLEGCKTKYVPELPL 116  
 146 NTSIIIPHNESGSLRTVHSVLRSPPELVAELVLDVDFSDREHLKPLEDYW-ALFP 204  
 117 TTSVIVFHNAMSTLRTVHSVLRSPPELVAELVLDVDFSDREHLKPLEDYW-ALFP 176  
 205 SVRLRTKREGLRTVRLGASVATGDIYTFELDSCEANVMPLDRIANRRTVCP 264  
 177 PVHVRMEQKSLRLKGAASVQVITFLDAECCTGWLPLARIKHDKRTVCP 236  
 265 MIDVIDHDFEYETQAG-DAMRGAPEMMEYKRIPIPE---ELQADSPDFESPVMAG 320  
 237 IIVDISDTFEY--MAGSDMTYGGNMTLNFRWYVPREMDRKRGGDTTLVPTTAMG 294  
 321 LPAVDKRMFWELGYPGLLEIWSGQVYISFKVMCGRMEDIPCSRVGHYIKYVYKV 380  
 295 LFSIDRDFQFIRGTADAMDIWGGENLEISFRIMCGGTLLEIVTCSHGVRKATPYTF 354  
 381 PAGVS--LARKLVAVYMEDEYAYTORREYVHLSAGVAVQKLRSSLNCSFKPF 438  
 355 PGGTGOIINKNRRLAEVWMBEFTFTTISPGVTKYVYGISSLRHKLQCKPFSWY 414  
 439 MTKIAMD--LRFYPPVPPAAAGELRNVTGTLCAPIKHALGSLPLEGCVARGSA 496  
 415 LENIYPSQIRRH-----SLGEINVTETNOCLDMARKENKVGITFNC-HMG--- 463  
 497 WNNQVFTFWREDIRDPQHTKKCFDAISHTSPVTLTYDCHSNKQMLKRYRDK-TL 555  
 464 --GNQVFESYANKERITD-----LCIDVSKLNPVMTLCKHLKQNLWEYDVKLT 515  
 556 YHPVSGSMD-CSESDHRI-FMNTGNPSLNOQMFEHTNSTVLEKF 600  
 516 QVANSNOCLDKATEEDSQVPSIRDCS-GSRSQQWLLR--NVTLPPIIF 559  
 RESULT 19  
 ID\_GLI1\_RAT STANDARD: PRT; 559 AA.  
 AC Q10473;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Polypeptide N-acetylglucosaminyltransferase 1 (EC 2.4.1.41)  
 DE (Protein-UDP acetylglucosaminyltransferase 1) (UDP-  
 DE GalNAc:polypeptide N-acetylglucosaminyltransferase 1) (Polypeptide  
 DE GalNAc:transferase 1) (GalNAc-T1) (p-GalTase 1) [Contains:  
 DE Polypeptide N-acetylglucosaminyltransferase 1, soluble form].  
 GN Name=GalnT1;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Sublingual gland;  
 RX MEDLINE=96318029; PubMed=8748168;  
 RA Hagen F., Gregorie C.A., Tabak L.A.;  
 RT "Cloning and sequence homology of a rat UDP-GalNAc:polypeptide N-

RT acetylglucosaminyltransferase.";  
 RL Glycoconj. J. 12:901-909(1995).  
 RN [2]  
 RP MUTAGENESIS OF CYS-106; CYS-212; CYS-214; CYS-235; CYS-330; CYS-339  
 RP AND CYS-408.  
 RX PubMed=1219709;  
 RA Tanno M., Toba S., Kozdy F.J., Elhammer A.P., Kurosaka A.;  
 RT "Identification of two cysteine residues involved in the binding of  
 RT UDP-GalNAc to UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase  
 RT 1 (GalNAc-T1)."  
 RL Eur. J. Biochem. 269:4308-4316(2002).  
 RN [3]  
 RP MUTAGENESIS OF CYS-442; CYS-459; CYS-482; CYS-497; CYS-523; CYS-540;  
 RP ASP-444; GLY-455; PHE-457; PHE-468; ASN-465; GLN-466; ASP-484 AND  
 RP ASP-525.  
 RX PubMed=12364335; DOI=10.1074/jbc.M207369200;  
 RA Tanno M., Saeki A., Kozdy F.J., Elhammer A.P., Kurosaka A.;  
 RT "The lectin domain of UDP-GalNAc:polypeptide N-  
 RT acetylglucosaminyltransferase 1 is involved in O-glycosylation of a  
 RT polypeptide with multiple acceptor sites."  
 RL J. Biol. Chem. 277:47088-47096(2002).  
 RN [4]  
 RP MUTAGENESIS OF ASP-444; ASP-484 AND ASP-525.  
 RX PubMed=12419318;  
 RA Tanno M., Kozdy F.J., Elhammer A.P., Kurosaka A.;  
 RT "Function of the lectin domain of polypeptide N-  
 RT acetylglucosaminyltransferase 1."  
 RL Biochem. Biophys. Res. Commun. 298:755-759(2002).  
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has a broad spectrum of substrates for peptides  
 CC such as EA2, Muc5AC, Muc6, Muc7, Muc8 and Muc9.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium.  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi; resides  
 CC evenly across the Golgi stack. A secreted form also exists (By  
 CC similarity).  
 CC -1- TISSUE SPECIFICITY: Heart, brain, spleen, liver, skeletal muscle  
 CC and kidney.  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called G1 motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC Gal/GalNAc-T motif), which is probably involved in catalytic  
 CC reaction and UDP-Gal binding.  
 CC -1- DOMAIN: The lectin B-type lectin domain directs the glycopeptide  
 CC specificity. It is required in the glycopeptide specificity of  
 CC enzyme activity but not for activity with naked peptide  
 CC substrates, suggesting that it triggers the catalytic domain to  
 CC act on GalNAc-glycopeptide substrates.  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
 CC subfamily.  
 CC -1- SIMILARITY: Contains 1 lectin B-type lectin domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC EMBL, U35890; AAC52511.1; -.  
 CC RCD, 620358; GalnT1.  
 DR InterPro: IPR001173; Glyco trans. 2.  
 DR InterPro: IPR008997; Rictin\_1like.  
 DR InterPro: IPR000772; Rictin\_B\_lectin.  
 DR Pfam: PF00535; Glycosyltransferase\_2; 1.  
 DR Pfam: PF00652; Rictin\_B\_lectin; 3.  
 DR SMART, SM00458; RICTIN; 1.

DR PROSITE, PS50231; RICIN B LECTIN; 1.  
 KM Calcium; Glycoprotein; Glycosyltransferase; Golgi stack; Lectin;  
 KM Manganese; Signal-anchor; Transmembrane.  
 FT PROPEP 1 40 Removed in soluble polypeptide N-acetylglactosaminyltransferase.  
 FT CHAIN 41 559 acetylglactosaminyltransferase 1, soluble form.  
 FT DOMAIN 1 8 Cytoplasmic (potential).  
 FT TRANSSEM 9 28 Signal-anchor for type II membrane protein (potential).  
 FT DOMAIN 29 559 Luminal, catalytic (potential).  
 FT DOMAIN 115 225 Catalytic subdomain A.  
 FT DOMAIN 285 347 Catalytic subdomain B.  
 FT DOMAIN 429 551 Ricin B-type lectin.  
 FT SITE 141 141 Not glycosylated (by similarity).  
 FT DISULFID 442 459 By similarity.  
 FT DISULFID 482 497 By similarity.  
 FT DISULFID 523 540 By similarity.  
 FT CARBOHYD 95 95 N-linked (GlcNAc...) (by similarity).  
 FT CARBOHYD 117 117 O-linked (potential).  
 FT CARBOHYD 118 118 O-linked (potential).  
 FT CARBOHYD 119 119 O-linked (potential).  
 FT CARBOHYD 288 288 O-linked (potential).  
 FT CARBOHYD 552 552 N-linked (GlcNAc...) (by similarity).  
 FT CARBOHYD 106 106 C->A: Loss of function.  
 FT MUTAGEN 212 212 C->A: Loss of function due to absence of interaction between UDP moiety and UDP-GalNAc.  
 FT MUTAGEN 214 214 C->A: Loss of function due to absence of interaction between UDP moiety and UDP-GalNAc.  
 FT MUTAGEN 235 235 C->A: No effect.  
 FT MUTAGEN 330 330 C->A: Loss of function.  
 FT MUTAGEN 339 339 C->A: Loss of function.  
 FT MUTAGEN 408 408 C->A: Loss of function.  
 FT MUTAGEN 442 442 C->A: Loss of function.  
 FT MUTAGEN 444 444 D->A: Induces a strong decrease in activity; Loss of function; when associated with A-484 and A-525.  
 FT MUTAGEN 455 455 F->Q: Induces a decrease in activity.  
 FT MUTAGEN 457 457 F->A: Little or no effect.  
 FT MUTAGEN 459 459 C->A: Loss of function.  
 FT MUTAGEN 465 465 N->A: Little or no effect.  
 FT MUTAGEN 466 466 Q->A: Induces a decrease in activity.  
 FT MUTAGEN 468 468 F->A: Little or no effect.  
 FT MUTAGEN 482 482 C->A: Loss of function.  
 FT MUTAGEN 484 484 D->A: Loss of function; when associated with A-444 and A-525.  
 FT MUTAGEN 497 497 C->A: Loss of function.  
 FT MUTAGEN 523 523 C->A: Loss of function.  
 FT MUTAGEN 525 525 D->A: Loss of function; when associated with A-444 and A-484.  
 FT MUTAGEN 540 540 C->A: Loss of function.  
 FT SEQUENCE 559 AA; 64229 MM; SE36A95D9422C833 CRC64;  
 Query Match 34.0%; Score 1116; DB 1; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 1,1e-81;  
 Matches 230; Conservative 86; Mismatches 171; Indels 38; Gaps 16;

QY 265 MIDVIDHDDPRVETQAG-DAMRGAFDWMYTKRIPPP---ELQKADSPDFESPVMAG 320  
 DB 237 IIDVIDSDTEY--MAGSDMTYGGFWKKNFRTVPVQREMDRKGRGRTIPVPTMAG 294  
 QY 321 LFAVDKRMFELGYDGLIWGBOYEISPKYMMCGRMEDIPCSFVGH1YRKRYVKV 380  
 DB 295 LFSIDRNYFQELIGYDAGMDIWGENLEISFRIWQCGTLEIYVCSHVGHVFRATPYTF 354  
 QY 381 PAGVS--LAARKVAVVWDEVALEYTYORPREYRHLSAGDVAQKLRSLNCKSKWF 438  
 DB 355 PGGGQIINKNGLAEVWMDDEFNFFYIISPGVTKYDYDISRVGLRHKDQCKPSWY 414  
 QY 439 MTKIAMD--LPRKPPVPEPAAAMGEIRNVTGGLCAVTKGALGSPRLSGCVGRBAA 496  
 DB 415 LENTYPSQIPRHY-----FSLGEIRNVTNOCCLDNMARKENKYGIRNC-HGMG--- 463  
 QY 497 WNNQVETPTWRDIRPDPOHTKRCFEDAISHTSPVTLVYCHSMKNQOLMYRKDK-TL 555  
 DB 464 --GNQVSYTANKEIRDD-----LCIDVSKLNGPVTMLKCHHLKGNQLMVEYDPAVLT 515  
 QY 556 YHPVSGSCMD-CBSRDHRT-FNATCNPSLSLTQWLFPHNYSVLEKP 600  
 DB 516 QHVNINQCLDKATEBDSQVPSIRDCT-GSRSQOMLR--NVTLPK 559  
 RESULT 20  
 GLTI\_BOVIN STANDARD; PRT; 559 AA.  
 ID GLTI\_BOVIN  
 AC 007537;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Polypeptide N-acetylglactosaminyltransferase 1 (EC 2.4.1.11)  
 DE (Protein-UDP acetylglactosaminyltransferase 1) (UDP-  
 DE GalNAc:polypeptide N-acetylglactosaminyltransferase 1) (Polypeptide  
 DE GalNAc transferase 1) [GalNAc-T1] (pp-GalNAc 1) [Contains:  
 DE Polypeptide N-acetylglactosaminyltransferase 1, soluble form].  
 GN Name=GLNT1;  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 41-74.  
 RC TISSUE=Colostrum, and Intestine;  
 RX MEDLINE=93286099; PubMed=7685345;  
 RA Homa F.L., Hollander T., Lehman D.J., Thomson D.R., Elhammer A.P.;  
 RT "Isolation and expression of a cDNA clone encoding a bovine UDP-  
 RT GalNAc:polypeptide N-acetylglactosaminyltransferase.";  
 RL J. Biol. Chem. 268:12609-12616 (1993).  
 RN [2]  
 RP SEQUENCE OF 41-559 FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Colostrum, and Placenta;  
 RX MEDLINE=93366815; PubMed=8360184;  
 RA Hagen F.K., van Wuyckhuysen B., Tabak L.A.;  
 RT "Purification, cloning, and expression of a bovine UDP-GalNAc:  
 RT polypeptide N-acetylglactosaminyltransferase.";  
 RL J. Biol. Chem. 268:18960-18965 (1993).  
 RN [3]  
 RP MUTAGENESIS OF ASN-95, HIS-125, HIS-137; ASN-141, HIS-146; HIS-179;  
 RP HIS-211; HIS-228; HIS-341; HIS-344; HIS-404; HIS-427; HIS-460;  
 RP HIS-498; HIS-499; HIS-517 AND ASN-552.  
 RX PubMed=9359852;  
 RA Wraag S., Hagen F.K., Tabak L.A.;  
 RT "Identification of essential histidine residues in UDP-N-acetyl-D-  
 RT galactosamine:polypeptide N-acetylglactosaminyltransferase-T1.";  
 RL Biochem. J. 328:193-197 (1997).  
 CC -!- FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has a broad spectrum of substrates for peptides  
 CC such as E2, Muc5AC, Muc1a, Muc1b and Muc7.



CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium.  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi; resides  
 CC evenly across the Golgi stack. A secreted form also exists (By  
 CC similarity).  
 CC -1- TISSUE SPECIFICITY: Colostrum contains a soluble form.  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called GT1 motif),  
 CC which is probably involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC GalINac-T motif), which is probably involved in catalytic  
 CC reaction and UDP-Gal binding (By similarity).  
 CC -1- DOMAIN: The ricin B-type lectin domain directs the glycopeptide  
 CC specificity. It is required in the glycopeptide specificity of  
 CC enzyme activity but not for activity with naked peptide  
 CC substrates, suggesting that it triggers the catalytic domain to  
 CC act on GalINac-glycopeptide substrates (By similarity).  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalINac-T  
 CC subfamily.  
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL: L07780; AAA30532.1; -;  
 CC EMBL: L17437; AAA68489.1; -;  
 CC PIR: A45987; A45987.  
 CC InterPro: IPR001173; Glyco trans\_2.  
 CC InterPro: IPR008997; RicinB-like.  
 CC InterPro: IPR000772; RicinB\_lectin.  
 CC Pfam: PF00535; Glycos transf\_2; 1.  
 CC Pfam: PF00652; RicinB\_lectin; 3.  
 CC SMART: SM00458; RICIN\_B\_lectin; 1.  
 CC PROSITE: PS50231; RICIN\_B\_LECTIN; 1.  
 CC K1: Calcium; Direct protein sequencing; Glycosyltransferase;  
 CC K2: Golgi stack; Lectin; Manganese; Signal-anchor; Transferase;  
 CC K3: Transmembrane.  
 CC PROPEP 1 40 Removed in soluble polypeptide N-  
 CC FT CHAIN 41 559 acetyl-galactosaminyltransferase.  
 CC FT DOMAIN 1 8 polypeptide N-  
 CC FT TRANSMEM 9 28 acetyl-galactosaminyltransferase 1,  
 CC FT soluble form.  
 CC FT CYTOPLASMIC (Potential).  
 CC FT SIGNAL-ANCHOR FOR TYPE II MEMBRANE  
 CC FT protein (Potential).  
 CC FT CATALYTIC SUBDOMAIN A.  
 CC FT CATALYTIC SUBDOMAIN B.  
 CC FT NOT GLYCOSYLATED (Probable).  
 CC FT BY SIMILARITY.  
 CC FT BY SIMILARITY.  
 CC FT BY SIMILARITY.  
 CC FT N-LINKED (GLCNAC: . .) (Probable).  
 CC FT O-LINKED (Potential).  
 CC FT O-LINKED (Potential).  
 CC FT O-LINKED (Potential).  
 CC FT O-LINKED (Potential).  
 CC FT N-LINKED (GLCNAC: . .) (Probable).  
 CC FT N-Q: Induces decrease in glycosylation.  
 CC FT H-A: Induces a strong decrease in  
 CC activity.  
 CC FT H-A: No effect.  
 CC FT H-Q: No effect.  
 CC FT H-A: No effect.  
 CC FT H-A: No effect.  
 CC FT H-A: No effect.  
 CC FT H-A: Loss of function.  
 CC FT MUTAGEN 137 137  
 CC FT MUTAGEN 141 141  
 CC FT MUTAGEN 146 146  
 CC FT MUTAGEN 179 179  
 CC FT MUTAGEN 211 211

FT MUTAGEN 228 228 H-A: No effect.  
 FT MUTAGEN 341 341 H-A: Induces a strong decrease in  
 FT activity.  
 FT MUTAGEN 344 344 H-A: Loss of function.  
 FT MUTAGEN 404 404 H-A: No effect.  
 FT MUTAGEN 427 427 H-A: No effect.  
 FT MUTAGEN 460 460 H-A: No effect.  
 FT MUTAGEN 498 498 H-A: No effect.  
 FT MUTAGEN 499 499 H-A: No effect.  
 FT MUTAGEN 517 517 H-A: No effect.  
 FT MUTAGEN 552 552 N-Q: Induces decrease in glycosylation.  
 FT SEQUENCE 559 AA; 64192 MW; B3E538C4D5E69B40 CRC64;  
 SO  
 Query Match 34.0%; Score 1115; DB 1; Length 559;  
 Best Local Similarity 43.6%; Pred. No. 1,4e-81;  
 Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;  
 QY 88 GNGEQGRP--YPMTDARVDQANRENGFNIVSDKISLNSLPPIRHPNCKSKRYLETLP 145  
 DB 57 GPGEMGRPVVLPKEDQKMKEMPKINQFNLMASMIANSLPVRLEGCTKYVPPNLP 116  
 QY 146 NTSIIIFHNGGWSLIRTVSHVLRSPPELVAVIYVDPSPDRHLKKPLEDYM-ALFP 204  
 DB 117 TTSVIVIFHNEAMSTLIRTVSHVLRSPPELVAVIYVDPSPDRFLKRPLESTYKLV 176  
 QY 205 SVRLRTKRGEGRLIRTMLGASVATGDTITFLDSHCANVWMLPPLDRIARNKITYCP 264  
 DB 177 PVHVRMEGRGGLIRALKGAAGVQVITFLDHCCTGWMLEPLARIKHDRKTVCP 236  
 QY 265 MIDVIDHDDPRFETQAG-DAMRGAFDMEYTKRPIPP--ELQKAPSPDFEPVWAGG 320  
 DB 237 IIDVISDITFEY--MAQSDMTYGGFNWKNFRWYVPVQREWRKGRGTLVPRTPWAGG 294  
 QY 321 LPAVDRKFMELGSDYDGLBETWGEQYEISPKVMCGRMEDIPCSRGHVIRKYVPEYKV 380  
 DB 295 LFSIDRDFQETIGYDAGMDTWGSENEISFRIQCGTLEIYCSHVGHFRATYTF 354  
 QY 381 PAGVS--LARNLKRAVWDEVAEYIYQRRPEYRLSAGDAVAQKLRSSLNCKSPKWF 438  
 DB 355 PGGTGQILINKNNRRLAEVWDEFNKPFYIISPGVTKYDGDISRLGLRHLQCRPSWY 414  
 QY 439 MTKLAMP-LPKFPVPPPPAANGELIRNVGTGCLADTKIGALGSLRLBECVNGREAA 496  
 DB 415 LENIYPPSQIPRHY-----FSIGBIRNVETNOCCLNMARKEKXGIFNC-HGMG--- 463  
 QY 497 WNNQVFTPTWRERDIPRDPDHTKKCFDAISHTSPVLYDCHSMKGNOLMKYRKDK-TL 555  
 DB 464 --GQVVSSTYNNKEIRTD-----LCDVSKLNGPVTMLCKHLLKGNQMLEYPAVLTL 515  
 QY 556 YHPVSGSCMD-CSESDHRI-FMNTCNPSLTQWLFHTNSTYLEKF 600  
 DB 516 QHVNNSNCLDKATDSDQVPSIRDCS-GSRQQLLR--NVTLPRIIF 559  
 RESULT 21  
 GLTI\_MOUSE STANDARD; PRT; 559 AA.  
 AC 008912; OTTNDL;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 45, Last sequence update)  
 DT 01-OCT-2004 (Rel. 40, Last annotation update)  
 DE Polypeptide N-acetyl-galactosaminyltransferase 1 (EC 2.4.1.41)  
 DE (Protein-UDP acetyl-galactosaminyltransferase 1) (UDP-  
 DE GalINac:polypeptide N-acetyl-galactosaminyltransferase 1) (Polypeptide  
 DE GalINac transferase 1) (GalINac-T) (pp-GaNTase 1) [Contains:  
 DE Polypeptide N-acetyl-galactosaminyltransferase 1, soluble form].  
 GN Name=GalNT1;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxId=10090;  
 RN [1]  
 SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.



RX MEDLINE:97298094; PubMed:9153242;  
 RA Hagen F.K., Ten Hagen K.G., Beres T.M., Balys M.M.,  
 RA Vanmyckhuyse B.C., Tabak L.A.,  
 RT "cDNA cloning and expression of a novel UDP-N-acetyl-D-  
 RT galactosamine:polypeptide N-acetylglucosaminyltransferase.";  
 RT J. Biol. Chem. 272:13843-13848(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SRRATN:CS7BL/607; TISSUE=Embryo;  
 RX MEDLINE:22388257; PubMed:12477932; DOI=10.1073/pnas.242603899;  
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Schenker C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datchenko L., Marasini K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Wyers R.M.,  
 RA Butterfield Y.S.N., Krzywicki M.I., Skalek U., Smalins D.E.,  
 RA Schnerch A., Schein J.E., Jones S.U.M., Mitra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP DOMAIN, AND MUTAGENESIS OF HIS-125; GLU-127; GLU-150; ASP-155;  
 RP ASP-156; ASP-209; HIS-211; GLU-213; ASP-310; GLU-319; ASN-320;  
 RP GLU-322; HIS-341; ASP-375; GLU-376; ASP-444; GLY-455; PHE-457;  
 RP HIS-460; GLY-464; ASN-465 AND GLN-466.  
 RX PubMed:10037781;  
 RA Hagen F.K., Hazes B., Raffo R., desa D., Tabak L.A.,  
 RT "Structure-function analysis of the UDP-N-acetyl-D-  
 RT galactosamine:polypeptide N-acetylglucosaminyltransferase. Essential  
 RT residues lie in a predicted active site cleft resembling a lactose  
 RT repressor fold.";  
 RT J. Biol. Chem. 274:6797-6803(1999).  
 RN [4]  
 RP TISSUE SPECIFICITY  
 RX PubMed:12651884; DOI=10.1093/glycob/cwg062;  
 RA Young W.W. Jr., Holcomb D.R., Ten Hagen K.G., Tabak L.A.,  
 RT "Expression of UDP-GalNAc:polypeptide N-  
 RT acetylglucosaminyltransferase isoforms in murine tissues determined  
 RT by real-time PCR: a new view of a large family.";  
 RT Glycobiology 13:549-557(2003).  
 RL -1- FUNCTION: Catalyzes the initial reaction in O-linked  
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-  
 CC galactosamine residue to a serine or threonine residue on the  
 CC protein receptor. Has a broad spectrum of substrates for peptides  
 CC such as EA2, MUC5AC, Muc1a, Muc1b and Muc7.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =  
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.  
 CC -1- COFACTOR: Manganese and calcium (By similarity).  
 CC -1- PATHWAY: Glycosylation.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi; resides  
 CC evenly across the Golgi stack. A secreted form also exists (By  
 CC similarity).  
 CC -1- TISSUE SPECIFICITY: Widely expressed at high level. Higher  
 CC expression in kidney, heart, small intestine and cervix and to a  
 CC lesser extent in all the other tissues tested.  
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase  
 CC region: the N-terminal domain (domain A, also called Gm motif),  
 CC which is primarily involved in manganese coordination and substrate  
 CC binding and the C-terminal domain (domain B, also called  
 CC Gal/GalNAc-T motif), which is probably involved in catalytic  
 CC reaction and UDP-Gal binding.  
 CC -1- DOMAIN: The ricin B-type lectin domain directs the glycopeptide of  
 CC specificity. It is required in the glycopeptide specificity of

CC enzyme activity but not for activity with naked peptide  
 CC substrates, suggesting that it triggers the catalytic domain to  
 CC act on GalNAc-glycopeptide substrates (By similarity).  
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T  
 CC subfamily.  
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
 CC -----  
 CC EMBL: U73820; AAB58477.1; -;  
 CC EMBL: BC056215; AAB56215.1; -;  
 CC MGD: MGI:894693; Galn1.  
 CC GO: GO:0004653; F:polypeptide N-acetylglucosaminyltransferase. . . ; IDA.  
 CC GO: GO:0006493; P:O-linked glycosylation; IDA.  
 CC InterPro: IPR001173; GlycoTrans\_2.  
 CC InterPro: IPR008997; RicinB\_1like.  
 CC InterPro: IPR000772; RicinB\_1lectin.  
 CC Pfam: PF00535; GlycosTrans\_2; 1.  
 CC Pfam: PF00652; RicinB\_1lectin; 3.  
 CC SMART: SM00458; RICIN\_1.  
 CC PROSITE: PS50231; RICIN\_B\_LECTIN\_1.  
 CC Calcutt; Glycoprotein; Glycosyltransferase; Golgi stack; Lectin;  
 CC Manganese; Signal-anchor; Transmembrane.  
 CC PROPEP 1 40  
 CC CHAIN 41 559  
 CC -----  
 CC DOMAIN 1 9  
 CC TRANSMEM 10 28  
 CC -----  
 CC FT DOMAIN 29 559  
 CC FT 115 225  
 CC FT 285 347  
 CC FT 429 551  
 CC FT 442 459  
 CC FT DISULFID 482 497  
 CC FT 523 540  
 CC FT DISULFID 523 540  
 CC FT CARBOHYD 95 95  
 CC FT 117 117  
 CC FT CARBOHYD 118 118  
 CC FT 119 119  
 CC FT CARBOHYD 288 288  
 CC FT 552 552  
 CC FT 125 125  
 CC FT 125 125  
 CC FT NOTAGEN 127 127  
 CC FT 150 150  
 CC FT NOTAGEN 155 155  
 CC FT 156 156  
 CC FT NOTAGEN 209 209  
 CC FT 211 211  
 CC FT NOTAGEN 213 213  
 CC FT 310 310  
 CC FT NOTAGEN 319 319  
 CC FT 320 320  
 CC FT NOTAGEN 322 322  
 CC FT 341 341  
 CC FT NOTAGEN 375 375  
 CC FT 376 376  
 CC FT NOTAGEN 444 444  
 CC FT 455 455  
 CC FT NOTAGEN 457 457  
 CC FT 457 457  
 CC FT NOTAGEN 457 457  
 CC -----  
 CC E-Q: Loss of function.  
 CC D-N,A,E: Loss of function.  
 CC H-A,D: Loss of function.  
 CC E-Q: Loss of function.  
 CC D-N: Loss of function.  
 CC N-A: No effect.  
 CC E-Q: Loss of function.  
 CC H-A,L,V,K,R: Little or no effect.  
 CC E-Q: No effect.  
 CC D-H: Induces a decrease in activity.  
 CC G-V: Induces a slight decrease in activity.  
 CC F-H: Induces a slight decrease in activity.

```

FT MUTAGEN 460 460 H->K,N: Induces a slight decrease in
FT MUTAGEN 464 464 G->A: Induces a slight decrease in
FT MUTAGEN 465 466 NO->QN: Induces a decrease in activity.
FT MUTAGEN 465 465 N->A: Little or no effect; induces a
FT MUTAGEN 465 465 slight decrease in activity when
FT MUTAGEN 466 466 associated with A-466.
FT MUTAGEN 466 466 O->A: Little or no effect; induces a
FT MUTAGEN 466 466 slight decrease in activity when
FT MUTAGEN 466 466 associated with A-465.
FT MUTAGEN 466 466 I->T (in Ref. 2).
FT CONFLICT 81 81
FT SEQUENCE 559 AA; 64255 MW; ACSAB55D91P83E4 CRC64;
SQ SEQUENCE 559 AA; 64255 MW; ACSAB55D91P83E4 CRC64;

Query Match 34.0%; Score 1115; DB 1; Length 559;
Best Local Similarity 43.6%; Pred. NO. 1.4e-81;
Matches 230; Conservative 87; Mismatches 172; Indels 38; Gaps 16;

QY 88 GNGEGRD-YPMTDARVDQAYRENGFNIVSDKISLNSLPLDIRHPCNSKRYLETLP 145
DB 57 GPGEMGKRVIPKEDQEKEMFKINQENLMASEMIALNRSILPVRLEGCKTKVYPDLF 116
QY 146 NTSIIIPHNNGSSSLRTHTSVNLRSPPELVAVLVDDPSDRLKKPLIEDTM-ALFP 204
DB 117 TTSVIVFHNMAWSTLLRTVSHVNRSDPHMIEIVLVDASERDFLRPLESYVKLKV 176
QY 205 SVRIKTKREGSLRTRMLGASVATGVITFDHSCENVMWMLPLDIRANRRTIVCP 264
DB 177 PVAVIRKESGSLRARKGAASVNGOYITFLDACECTAGLELARKINHRTIVCP 236
QY 265 MIDVIDHDDFRYETQAG-DAMRGAFDWMYKRIPIRP--ELQKADSDPFESPVNAG 320
DB 237 IIDVIDSDTFEY--MAGSDMTYGGHNMKINFRWYVPREMDRKRKGRDTLPVRRPTWAG 294
QY 321 LFAVDKRFMLGCDPLGLTWGSGQVYSFKVMCGRMEDIPCSRGHYIKRVKPVKV 380
DB 295 LFSIDRDVFOEIRIGYDAGMDIWGGENLEISFRIMCGGTLTIVTCSHGHFRKATPTP 354
QY 381 PAGS--LARIKRAVWMDVAVIYQRRPEYHLGAGVAOVOKLRSLSCKSPFMF 438
DB 355 PGCGIOLINKNNRRLAEVWMDFKNFYIISBGTAKVDYGDSSRLGLRKLCKCPFWY 414
QY 439 MTKIAMD-LPKFYPPVPPAAWGEIRNVGTGLCADTRGALGSPFLREGCVRGEEAA 496
DB 415 LENIYPSQIDPRHY-----FSLGEIRNVETNQLDMARKENKVGIFNC-HGMG--- 463
QY 497 WNNNOVFPTWRBEDIIRPDPOHTKKFCDDAISHSPTVLYHCHSKMGKQWTKYRKDK-TL 555
DB 464 --GNQVSESYTANKELRTDD-----LCIDVSKLNGPVMLKCHLKGQMLEYDVPVXLTL 515
QY 556 YHPVSGCMD-CSESDBRI-FMNTGNPSLTQQWLFHTNSTVLEKF 600
DB 516 QHVNNSQCLDKATBEDSQVPSIRCT-GRSQQMLLR--NTLPEIF 559

```

```

RX MEDLINE=98192620; PubMed=9525933;
RA Hagen F.K., Nehke K.;
RT "cDNA cloning and expression of a family of UDP-N-acetyl-D-
RT galactosamine:polypeptide N-acetylgalactosaminyltransferase sequence
RT homologs from Caenorhabditis elegans."
RL J. Biol. Chem. 273:8268-8277(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Stoneking T., Wohlmann P., Lennox S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Probable glycopeptide transferase involved in O-linked
CC oligosaccharide biosynthesis. Glycopeptide transferases catalyze
CC the transfer of an N-acetyl-D-galactosamine residue to an already
CC glycosylated peptide (By similarity). In contrast to other members
CC of the family, it does not act as a peptide transferase that
CC transfers GalNAc onto serine or threonine residue on peptides that
CC have been tested. Some peptide transferase activity is however not
CC excluded, considering that its appropriate peptide substrate may
CC remain unidentified.
CC -1- COFACTOR: Manganese and calcium (By similarity).
CC -1- PATHWAY: Glycosylation.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
CC similarity).
CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase
CC region: the N-terminal domain (domain A, also called G1 motif),
CC which is probably involved in manganese coordination and substrate
CC binding and the C-terminal domain (domain B, also called
CC Gal/GalNAc-T motif), which is probably involved in catalytic
CC reaction and UDP-Gal binding (By similarity).
CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and
CC contributes to the glycopeptide specificity (By similarity).
CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
CC subfamily.
CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: AF031841; AAC16677.1; -.
CC DR EMBL: AC006674; AAF60620.1; -.
CC DR PIR: T42251, T42251.
CC DR Wormpep: Y46H3A.6; CE24309.
CC DR InterPro: IPR001173; Glyco_transf_2.
CC DR InterPro: IPR008997; RicinB_1like.
CC DR InterPro: IPR000772; RicinB_lectin.
CC DR Pfam: PF00535; Glycos_transf_2; 1.
CC DR Pfam: PF00652; RicinB_lectin; 3.
CC DR SMART: SM00458; RICIN_1.
CC DR PROSITE: PS50231; RICIN_B_LECTIN; 1.
CC DR Calcium; Glycosyltransferase; Golgi stack; Lectin; Manganese;
CC Signal-anchor; Transferase; Transmembrane.
CC KW DOMAIN 1 20 Cytoplasmic (Potential).
CC FT TRANSMEM 21 38 Signal-anchor for type II membrane
CC FT DOMAIN 39 601 protein (Potential).
CC FT DOMAIN 155 265 Luminal (Potential).
CC FT DOMAIN 328 390 Catalytic subdomain A.
CC FT DOMAIN 477 595 Ricin B-type lectin.
CC FT DISULFID 490 506 By similarity.
CC FT DISULFID 529 542 By similarity.
CC FT DISULFID 568 583 By similarity.
CC FT CARBOHYD 135 135 N-linked (GlcNAc...) (Potential).
CC SQ SEQUENCE 601 AA; 68108 MW; 1A7E744C93BDCD57 CRC64;

Query Match 34.0%; Score 1113; DB 1; Length 601;
Best Local Similarity 42.0%; Pred. NO. 2.2e-81;
Matches 226; Conservative 90; Mismatches 186; Indels 36; Gaps 13;

```

```

QY 68 GQKLDHDKAIRDQVRVNGEGRPYPMTDAERV---DQAYRNGFNITYSDKISLN 124
DB 79 GSEIGNEYEPKPELPSNQ---PGEHGKVPVPTDEGMAAGAAEEFENFTYSDMTSMN 135
QY 125 RSLPDIRHPNCSKRYLETLPNTSIIIPHNMGSSLLRTVHSVLRSPPELVAEIVLD 184
DB 136 RTIIDIRECKHMDYPERKLPVSVVVVFHNEGTPPLRTVHSVLRSPPELIEQVWVD 195
QY 185 DFSREHLKKPLEDMALF-PSVRLRTKKREGLIRTMIGASVATGDTVITFLDSHCAN 243
DB 196 DSDSKPHLEKLDKDYTRFNGKVIYVTRTEGREGLINASIGAKISTGSEVFLDAHCEVN 255
QY 244 VNMPLPILLDIRIARRKTIIVCPMIDVIDHDDFRYEIQAGDA---MKGADNEMTYKRLPI 300
DB 256 TNNLPLPLAIPKRRKKTWTVPVIGDIDSNSWEYSVSPNHHSGIFEMGLKYEQTIT 315
QY 301 PE--LOKADSPDPESPVMAGGLFVADRKFMEGLGYDGLMEIGEGYEISFKVMWCG 358
DB 316 ERETNRKHNQSPRSPTHAGGLPAINLMKEGLGYBGLQINGEGYEISFKIMCGG 375
QY 359 RMEDIPCSRVGHITRYKVPY---KVPAGVSLARMLKRVAEVMDYAEYIVOREPEYHL 415
DB 376 GIVFVPCSHGVHVRSHMPYSFGKFSKRPVISINMMRVVXTWMDYSKYLTREPOATNV 435
QY 416 SAGDVAVOKLRSSLNCSFKMFMTKIAMDLPKPYPRPPRAANGELRANVTGICADTK 475
DB 436 NPGDISAQALARDKIQCKSFMYMENVAYDVAKSPML-PRPDVWGERNRNATGICLD-R 493
QY 476 HGAIGSPLELEGCVRGGEAAMNNQVFTFTWREDIRPGDPOHTKFCFPAISHTSPVTL 535
DB 494 MGGIIPGPGATGC-----HGYGNGQLRLNVQGMAGS-----WGLTA---NGIRI 537
QY 536 YDCHSMKG--NQLMKY-RKDKTYHPVSGSCMDCSDHRIEMNTCNPSLTLQOMLFE 590
DB 538 QANHCVKGTVNGFMSYDKTKQIHSQKRCITVSESESVTLQCTEDNRQKQFVK 595

```

## RESULT 23

Q6GM51 PRELIMINARY; PRT; 556 AA.

```

AC Q6GM51;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Butelostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OC NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RX MEDLINE=2338257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stropienko M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carrinci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullaly S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Boulford G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.R.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human

```

```

RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.,
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative."
RT Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Klein S., Strausberg R.,
RL Submitted (Jun-2004) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC074234; AAH74234.1;
DR InterPro; IPR001173; Glyco trans. 2.
DR InterPro; IPR008997; RicinB like.
DR InterPro; IPR00772; Ricin_B_lectin.
DR Pfam; PF00535; Glycof. transF_2;
DR Pfam; PF00652; Ricin_B_lectin; 3.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
KW Hypothetical protein.
SQ SEQUENCE 556 AA; 63908 MW; 3A9FF34A39AB655B CRC64;

```

```

Query Match 33.9%; Score 1112.5; DB 2; Length 556;
Best Local Similarity 42.8%; Pred. No. 2.2e-81;
Matches 234; Conservative 91; Mismatches 181; Indels 41; Gaps 16;

```

```

QY 64 FLGGQKLDHDKAIRDQVRVNGEGRPYPMTDAERV---DQAYRNGFNITYSDKISLN 116
DB 27 YFSECNCDKRSKSLPALRAVIRSPEGGEGLKAVIIPKDOQKKEKLPKINQFLM 86
QY 117 VSDKISLNRSLPDIRHNCNCKRYLETLPNTSIIIPHNMGSSLLRTVHSVLRSPPEL 176
DB 87 ASDLIANRSLPDIRHNCNCKRYLETLPNTSIIIPHNMGSSLLRTVHSVLRSPPEL 146
QY 177 VAEIVLVDDPSDRHKKPLEDYMA-LFPSVRLRTKKRGLIRTMIGASVATGDTVITF 235
DB 147 ISEIIVDDASERFLTPTPLENTYKHEVAVKILRMGORSGLIRARISGANVAAGKITF 206
QY 236 LDHSGEANVMPLPDLRIARNKRTIYCPMIDVIDHDDFRYEIQAG-DANRGAPEWEMY 294
DB 207 LDAHCECTFGWLBELRLRIKEDRKTVVCPIIDVISDTFXY--MAGSDMTYGGFNMKINF 264
QY 295 KRIRIPP---ELOKADSDPESPVMAGGLFVADRKFMEGLGYDGLMEIGEGYEISF 351
DB 265 RMYVPVPRMDRRKGDRTLPVPTPTMAGGLPSIDKTYFELGTVDSCMDIWMGENSEWF 324
QY 352 KVMNCGRMEDIPCSRVGHITRYKVPYKVPAGVS--LARNLKRVAEYVMDYAEYIVQR 409
DB 325 RIMQCGSLLEIVCSHGVRFRKATPYTFPGTGTHVINKNNRLAEVMDPFOFPYIIS 384
QY 410 PEYHLSAGDVAVOKLRSSLNCSFKMFMTKIAMDLPKPYPRPPRA-AMGEINNVG 467
DB 385 PGVVKLVYGVDSSEKALRENIKCPFGMYLETV-----YPSQIPRRYPSLSEIRNVE 437
QY 468 TGLCADPKHAGLSPLLEGCVRGGEAAMNNQVFTFTWREDIRPGDPOHTKFCFPAI 527
DB 438 TNOGLDMKREKREKVEGIFNC-HQMG-----GNQVFTYADKEIKRTD-----LCIDVS 485
QY 528 SHTSPVTLTYDCHSKKGNQLMKYRKDK-TLYHPVSGSCMDCSDHRIF--MNTCNPSSLT 584
DB 486 RINSPVIMLCKHNRGNQLMVEYDAEHLILRHINSNGLDDEADDDKVVPTIKECN-GSRS 544
QY 585 QQWLFEH 591
DB 545 QQWLIRN 551

```

## RESULT 24

AA056702 PRELIMINARY; PRT; 617 AA.  
 ID AA056702  
 AC AA056702  
 DT 02-MAR-2004 (Tremblrel. 27, Created)  
 DT 02-MAR-2004 (Tremblrel. 27, Last sequence update)  
 DT 02-MAR-2004 (Tremblrel. 27, Last annotation update)  
 DE UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41).  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 EX PubMed:12829714;  
 RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.;  
 RT "Functional Characterization and Expression Analysis of Members of the UDP-GalNAc:Polypeptide N-Acetylglucosaminyltransferase Family from Drosophila melanogaster";  
 RT J. Biol. Chem. 278:35039-35048(2003).  
 RL EMBL; AY268066; AA056702.1--  
 KW Glycosyltransferase; transferase.  
 SQ SEQUENCE 617 AA; 70482 MW; CA0A7A7A53B73E3 CRC64;  
 Query Match 33.9%; Score 1111.5; DB 2; Length 617;  
 Best Local Similarity 41.9%; Pred. No. 3e-81;  
 Matches 243; Conservative 82; Mismatches 194; Indels 61; Gaps 18;  
 QY 33 ERQDGTGGGGAAVAPAGGSGSHRQKRTFLGDKQ-----LKDMDKEAIRDQAQ 85  
 DB 69 EKSLDSESG-----ALIMGQ-----FASGISMTPSYVLKKMFLAPSVQEKAG 114  
 QY 86 RVNGGEQGR--YPMTDARDOAYRENGFNITYSDKLSLRSLPDHHPNCSRYLET 143  
 DB 115 K--FGEMCKPKVPIADMMDLKKKEKQFNLLASDMSLRSLTDVHECCRRGHVASK 172  
 QY 144 LPTNIIIPFNEGSSLLRTVSHVSNRSPPELVAEIYLVDFSPREHLKKPLDYMLF 203  
 DB 173 LPTTSIVFNEAWTILRTVSNRSPALKEIILVDDASERDPLGQLEIYAKL 232  
 QY 204 P-SVRIATKREGILRTMLGASVATGVTTFDLSHCANVMMLPLLDRIARRKTIIV 262  
 DB 233 PVKTFVARTKRSGLIRARLLGAHVSGEVITFLDAHCEGMEPLARIYQNRRTIV 292  
 QY 263 CPMLDIVDHDPRFETQGDANRGAFFWEMTKRIPIPE---LQADSPSPESPVMAG 319  
 DB 293 CPIIDVISDEFEEYIT-ASDSTWGGFNMKLNFRWYRVRSREMARNNDRTPALRTPTWAG 351  
 QY 320 GLPFAVDRKMFELGQYDGLGEIWEQGEYEIFKVMCGARMEDIPCNRVGHYIRKYVPYK 379  
 DB 352 GLFESIDKDYFEISYDEGMIDWGENLEMSFRIOCGGILIEIPCSVGHVFRDKSPYT 411  
 QY 380 VPAGVS--LARNLRVAEVMDEVAEYIYQRRPEYRLSADVAVQKLRSLNCKSPFK 437  
 DB 412 FPGGVAKIVLHNAARVAEVMDEWDFYMSSTGARSKASAGVSDRKALRLRDKSPFM 471  
 QY 438 FMTKIAMLPRFYVPEPPAAM--GEIRANGTGLCATYKAGALSPRLREGCYRGRSEA 495  
 DB 472 YLENV-----YPSLMLPLDYIYIGIRNAETETCLDTRKKNKKEGIYVC-HGLD-- 521  
 QY 496 AMNNQVETFTWRDIRPGDPQHTKKPCFDALISHTSPVLTVLDCSHMKNGQMLKY--RKQXT 554  
 DB 522 ---GNQVAYITKQOINSD-----LCLDASSSSGPNVNMVRCNMGANGQEWYDAEKM 572  
 QY 555 LYHPVSGSCMDCSDSHRIFMNT--CNPSSL--TOOQLFE 590  
 DB 573 IRTNMSGQCLQARATDD---ANTPLRLPCSYGKGQWLM 609

GL13 MOUSE STANDARD; PRT; 556 AA.  
 ID GL13 MOUSE  
 AC O8CF93; O8BLE4; O8BYT3;  
 DT 01-OCT-2004 (Rei. 45, Created)  
 DT 01-OCT-2004 (Rei. 45, Last sequence update)  
 DT 01-OCT-2004 (Rei. 45, Last annotation update)  
 DE Polypeptide N-acetylglucosaminyltransferase 13 (EC 2.4.1.41)  
 DE (Protein-UDP acetylglucosaminyltransferase 13) (UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase 13) (Polypeptide GalNAc transferase 13) (GalNAc-T13) (pp-GalTase 13).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND TISSUE SPECIFICITY.  
 RC TISSUE=Brain;  
 RX MEDLINE=22393469; PubMed=12407114; DOI=10.1074/jbc.M203094200;  
 RA Zhang Y., Iwasaki H., Wang H., Kudo T., Kalika T.B., Hennet T., Kubota T., Cheng L., Inaba N., Gotoh M., Togayachi A., Guo J.-M., Hiatsomi H., Nakajima K., Nishihara S., Nakamura M., Marth J.D., Naritatsu H.;  
 RT "Cloning and characterization of a new human UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylglucosaminyltransferase, designated pp-GalNAc-T13, that is specifically expressed in neurons and synthesizes GalNAc alpha-serine/threonine antigen.";  
 RT J. Biol. Chem. 278:573-584(2003).  
 RL [2]  
 RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina, and Hypothalamus;  
 RX MEDLINE=22354683; PubMed=12468851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Nakai I., Oseko N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T., Baldarelli R., Hill D.B., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Betsel K.W., Blake J.A., Bradt D., Brusic V., Chochia C., Coriani L.B., Cousins S., Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Fraser K.S., Gaasterland T., Gariboldi M., Giesl C., Godzik A., Gough J., Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis B.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Naashima T., Numa K., Okido T., Pavan W.J., Petosa G., Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S., Ravasi T., Reed J.C., Reed J.J., Reid J., Ring B.Z., Rindgen M., Sadelain L., Schneider C., Sempile C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Walestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arikawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yaunishi A., Yoshino M., Wacreston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [3]  
 RP FUNCTION.  
 RX PubMed=8618846;  
 RA Hennet T., Hagen F.K., Tabak L.A., Marth J.D.;  
 RT "T-cell-specific deletion of a polypeptide N-acetylglucosaminyltransferase gene by site-directed recombination.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:12070-12074(1995).  
 RN [4]  
 RP TISSUE SPECIFICITY.  
 RX PubMed=12651884; DOI=10.1093/glycob/CWG062;  
 RA Young W.W., Jr., Holcomb D.R., Ten Hagen K.G., Tabak L.A.;  
 RT "Expression of UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase isoforms in murine tissues determined

RESULT 25  
 GL13\_MOUSE

RT by real-time PCR: a new view of a large family.";

RL Glycobiology 13:549-557 (2003).

CC -1- FUNCTION: Catalyzes the initial reaction in O-linked oligosaccharide biosynthesis, the transfer of an N-acetyl-D-galactosamine residue to a serine or threonine residue on the protein receptor. Has a much stronger activity than GALNT1 to transfer GalNAc to mucin peptides, such as Muc5ac and Muc7. Able to glycosylate SDG3. Probably responsible for the synthesis of Tn antigen in neuronal cells.

CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide = UDP + N-acetyl-D-galactosaminyl-polypeptide.

CC -1- COFACTOR: Manganese and calcium (By similarity).

CC -1- PATHWAY: Glycosylation.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By similarity).

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named Isoforms=2;

CC Name=1;

CC IsoId=Q8CF93-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q8CF93-2; Sequence=VSP\_011220;

CC Note=No experimental confirmation available;

CC -1- TISSUE SPECIFICITY: Specifically expressed in neuronal cells. Not expressed in glial cells such as astrocytes. Expressed at low level.

CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase region: the N-terminal domain (domain A, also called GT1 motif), which is probably involved in manganese coordination and substrate binding and the C-terminal domain (domain B, also called Gal/galnac-t motif), which is probably involved in catalytic reaction and UDP-Gal binding (By similarity).

CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and contributes to the glycopeptide specificity (By similarity).

CC -1- MISCELLANEOUS: The absence of Galnt13 (which was wrongly assigned as Galnt8), does not induces clear defects in mice. It however abolishes Tn antigen in neuronal cells.

CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T subfamily.

CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AB082928; BACS45456.1; -

DR EMBL: AK038387; BAC29981.1; -

DR EMBL: AK045417; BAC32353.1; -

DR MGD: MGI:2139447; Galnt13.

DR InterPro: IPR001173; Glyco\_trans\_2.

DR InterPro: IPR008997; RicinB\_1like\_

DR InterPro: IPR000772; Ricin\_B\_lectin.

DR Pfam: PF00535; Glycos\_transf\_2; 1.

DR Pfam: PF00652; Ricin\_B\_lectin; 3.

DR SMART: SMO0458; RICIN; 1.

DR PROSITE: PS50231; RICIN\_B\_LECTIN; 1.

DR Alternative splicing: Calcium; Glycosyltransferase; Golgi stack;

KW Lectin; Manganese; Signal-anchor; Transferrase; Transmembrane.

KM Lectin; Manganese; Signal-anchor; Transferrase; Transmembrane.

FT DOMAIN 1 4

FT TRANSMEM 5 27

FT SIGNAL-ANCHOR (Potential)

FT CYTOPLASMIC (Potential)

FT PROTEIN-ANCHOR (Potential)

FT LUMENAL (Potential)

FT CATALYTIC SUBDOMAIN A.

FT CATALYTIC SUBDOMAIN B.

FT RICIN B-TYPE LECTIN.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT BY SIMILARITY.

FT CARBOHYD 94 94

FT CARBOHYD 116 116

FT N-LINKED (GlcNAc. . .) (Potential).

FT N-LINKED (GlcNAc. . .) (Potential).

FT CARBOHYD 551 551 N-linked (GlcNAc. . .) (Potential).

FT VARSPLIC 466 466 V -> VHDCLSPSLGVGAEBCNSNPLGLVYTPINRQ

FT FT (in isoform 2).

FT /FTId=VSP\_011220.

FT CONFLICT 457 457 N -> K (in Ref. 2; BAC32353).

FT SEQUENCE 556 AA; 63982 MW; 593934CPDOAED148 CRC64;

SO

Query Match 33.8%; Score 1109; DB 1; Length 556;

Best Local Similarity 43.8%; Pred. No. 4.1e-01;

Matches 227; Conservative 91; Mismatches 164; Indels 36; Gaps 15;

88 GNGEGRP--YPTDAERVDAAYRENGFNIVYSDKISLNSLPDIBHPNCSKRYETLP 145

56 GPEMKAVLIPIDDDQEMKELKINQFNLMAEDLIALNSLPDVLREGCKTYPELP 115

146 NTSIIIPFNEGSSILRTVSVLNSPPELVAEIVLVDDFSREHLKPLEDYM-ALFP 204

116 NTSVIVIFNHEASTLIRTVSVYNSPHYLSSEVILVDDASBRDFKLTLENVKTLV 175

205 SVRIIRTKREGILIRRMGLASVATGDTTFELDSHEANVMPLPILDRFARRKTYCP 264

176 PVKIRHEERSGLIRALRGAALSKQVITFLDAHCECTGLWLEPLIARIKEDKTYVCP 235

265 MIDVIDHDPRYETDAG-DAMRGAPDMENYKRIPIPP---ELQKADPSDFESPVAVAG 320

236 IIDVISDTEFY--MAGSDMTYGFNWKLFNRYPPVQRMEDRKDRKDRITLPVPTPMAG 293

321 LFAVDRKFWELGQYDPLEIWGEOYEISFKYVMCGRMEDIPCSRVGHIYRYVYKYV 380

294 LFSIDRNFEEIGTYDAGMDIWGEMLEMSFRIMCGGSLIEIVTCSHVGHVFRKATPYTF 353

381 PAGYS--LAANLRVAEVMDEYAEIYQRRPEYRHLNAGDAVAVOKLRSLNCKSRWF 438

354 PGSTGHVINKNNRRLAEVWMDERKDFPYIISPQVAKVDGVSVRKTLRENLLCKPESWY 413

439 MTXIAMDLPEKYPPVEPPAAW--GEIRNYGTGLCAPTRKHGALGSPRLREGVGRGEAA 496

414 LENI-----YFDSQIPRRTYSLGIRNVEVETNOCLDNMKRKENKYGIFNC-HGAG--- 462

497 WNNMOVFTPTWRBDIRPGDPQHTKKECFDAISHTSPVTLVDCHSMKGNQLMKYRKOK-TL 555

463 --GNQVSYTADKEIRTD-----LCLDVSRLSGPVMKCHHMGNGQMLEYDAERLTLL 514

556 YHPVSGSAMD-CESSDHRT-FNATCNPSLTQOMLEH 591

515 RHANSNQCIDEPSEBDMKVPYMODCS-GSRSQQWLRLN 551

Search completed: November 22, 2004, 13:43:40

Job time : 71 secs

**This Page Blank (uspto)**